Vol. IV 1933

# THE ANNALS

of

# MATHEMATICAL STATISTICS

Printed in U. S. A.



Published and Lithoprinted by
EDWARDS BROTHERS, INC.
ANN ARBOR, MICH.

# CONTENTS OF VOLUME IV

## CONTENTS OF VOLUME IV – Continued

*THE AMERICAN STATISTICAL ASSOCIATION*

Vol. IV No. 1

# THE ANNALS
of
# MATHEMATICAL STATISTICS

*SIX DOLLARS PER ANNUM*

FEBRUARY 1933

PUBLISHED QUARTERLY BY

AMERICAN STATISTICAL ASSOCIATION

*Publication Office*—Edwards Brothers, Inc., Ann Arbor, Michigan

*Business Office*—530 Commerce Bldg., New York Univ., New York, N. Y.

*Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.*

STATEMENT OF OWNERSHIP

UNDER ACT OF CONGRESS OF AUGUST 24, 1912

*Publisher*—American Statistical Association, New York City, New York.

*Editor*—H. C. Carver, University of Michigan.

*Managing Editor*—Rattan S. Seldon, Ann Arbor, Michigan.

*Business Manager*—J. W. Edwards, Ann Arbor, Michigan.

*Owner*—American Statistical Association, 530 Commerce Bldg., New York City.

# ON THE SYSTEM OF CURVES FOR WHICH THE METHOD OF MOMENTS IS THE BEST METHOD OF FITTING

by

A. L. O'Toole.

National Research Fellow.

In Mr. R. A. Fisher's paper[1] on the mathematical foundations of theoretical statistics the following statement is found: "The method of moments applied in fitting Pearsonian curves has an efficiency exceeding 80 per cent. only in the restricted region for which $\beta_2$ lies between the limits of 2.65 and 3.42 and for which $\beta_1$ does not exceed 0.1. It was, of course, to be expected that the first two moments would have 100 per cent. efficiency for the normal curve, for they happen to be the optimum statistics for fitting the normal curve. That the moment coefficients $\beta_1$ and $\beta_2$ also tend to 100 per cent. efficiency in this region suggests that in the immediate neighborhood of the normal curve the departures from normality specified by the Pearsonian formulas agree with those of that system of curves for which the method of moments gives the solution of the method of maximum likelihood.

The system of curves for which the method of moments is the best method of fitting may easily be deduced, for if the frequency in the range $dx$ be $y(x, \theta_1, \theta_2, \theta_3, \theta_4)dx$ then $\frac{\partial}{\partial \theta} \log y$ must involve $x$ only as polynomials up to the fourth degree; consequently

$$y = e^{-a^2(x^4 + p_1 x^3 + p_2 x^2 + p_3 x + p_4)}$$

[1] Philosophical Transactions of the Royal Society of London, vol. 222 series A (1921), p. 355.

the convergence of the probability integral requiring that the coefficient of $x^4$ should be negative, and the five quantities $a, p_1, p_2, p_3, p_4$ being connected by a single relation, representing the fact that the total probability is unity." It is with these curves having a fourth degree polynomial in the exponent that the present paper is concerned.

The first step in the study of this system of frequency functions is to find an expression for the value of the integral

$$I=\int_{-\infty}^{\infty} e^{-a^2(x^4+p_1x^3+p_2x^2+p_3x+p_4)}dx.$$

In other words, it is necessary to know how the integral depends on the parameters $a, p_1, p_2, p_3, p_4$.

Since $a$ depends only on the unit of measure of $x$ it will be sufficient for the moment to consider $a^2=1$. Furthermore a linear transformation on $x$ leaves the value of the integral unchanged. If we replace $x$ by $x-p_1/4$ the integral to be considered becomes

$$I=\int_{-\infty}^{\infty} e^{-(x^4+px^2+qx+r)}dx=k\int_{-\infty}^{\infty} e^{-(x^4+px^2+qx)}dx \quad \text{where } k=e^{-r}.$$

Consider now then the frequency curves $y=ke^{-(x^4+px^2+qx)}$. These curves are typically bimodal and may be classified according to the number and kind of modes. The positions of the modes are given by the solutions of the equation $\frac{dy}{dx}=0$, that is by the roots of the equation $4x^3+2px+q=0$. The discriminant of this cubic equation tells us that there will be three distinct real roots and thus two distinct maxima with a minimum between them for the curve, that is two distinct modes for the quartic exponential curve, if $-8p^3>27q^2$, $p<0$. Two roots will be real and equal if $-8p^3=27q^2$, $p<0$. The three roots will be real and equal if

$p = q = 0$, the three roots being $x = 0$. In the case of three real distinct roots, if two of the roots are equal in magnitude but opposite in sign then $q = 0$ and the curve is symmetrical with respect to the $y$-axis. If $p = 0, q \neq 0$ there will be one real root and two imaginary roots given by the three cube roots of $\frac{q}{4}$. That there will be a real maximum at the value of $x$ given by the real cube root of $\frac{q}{4}$ is easily seen from the nature of the curve or by considering points at values of $x$ on each side of this real cube root of $\frac{q}{4}$.

Hence the following classes of curves and their respective equations will be considered

Type I. $y = ke^{-x^4}$.

The curve which is symmetrical with respect to the $y$-axis and has only one mode, this mode being at $x = 0$.

Type II: $y = ke^{-(x^4 - 2bx^2)}$, $b > 0$.

The curve which is symmetrical with respect to the $y$-axis and has two distinct modes at $x = \pm\sqrt{b}$.

Type III: $y = ke^{-(x^4 - 4cx)}$, $c \neq 0$.

The asymmetrical curve with one real mode at $x = \sqrt[3]{c}$

Type IV: $y = ke^{-(x^4 + px^2 + qx)}$

The general type of curve with the quartic exponent.

Type I.

**First evaluate the definite integral**

$$I_0 = \int_{-\infty}^{\infty} e^{-x^4} dx = 2\int_0^{\infty} e^{-x^4} dx.$$

Let $x = y^{1/4}$, $dx = (1/4)y^{-3/4}dy$. Then Then

$$I_0 = 2(1/4)\int_0^\infty y^{-3/4}e^{-y}dy = (1/2)\int_0^\infty y^{1/4-1}e^{-y}dy = \frac{1}{2}\Gamma\left(\frac{1}{4}\right).$$

Similarly it may be shown that

$$\int_0^\infty x^p e^{-x^q}dx = \frac{1}{q}\Gamma\left(\frac{p+1}{q}\right), p > -1.$$

$I_1 = \int_{-\infty}^\infty x e^{-x^4}dx = 0$ since the integrand is an odd function.

$$I_2 = \int_{-\infty}^\infty x^2 e^{-x^4}dx = \frac{1}{2}\Gamma\left(\frac{3}{4}\right).$$

$$I_3 = \int_{-\infty}^\infty x^3 e^{-x^4}dx = 0.$$

$$I_4 = \int_{-\infty}^\infty x^4 e^{-x^4}dx = \frac{1}{2}\Gamma\left(\frac{5}{4}\right) = \frac{1}{8}\Gamma\left(\frac{1}{4}\right).$$

$$I_{2n-1} = \int_{-\infty}^\infty x^{2n-1}e^{-x^4}dx = 0, \quad n = 1, 2, 3, \ldots\ldots$$

$$I_{2n} = \int_{-\infty}^\infty x^{2n}e^{-x^4}dx = \frac{1}{2}\Gamma\left(\frac{2n+1}{4}\right), \quad n = 1, 2, 3, \ldots\ldots$$

Hence if the total frequency is unity then $k = \frac{2}{\Gamma(\frac{1}{4})}$.

$\mu_1 = I_1/I_0 = 0,$

$\mu_2 = I_2/I_0 = \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} = \frac{\sqrt{2\pi}}{B(\frac{1}{4},\frac{1}{4})} = 0.3379891$ approximately,

$\mu_3 = I_3/I_0 = 0,$

$\mu_4 = I_4/I_0 = 1/4,$

$\mu_{2n-1} = I_{2n-1}/I_0 = 0,$

$\mu_{2n} = I_{2n}/I_0 = \frac{\Gamma(\frac{2n+1}{4})}{\Gamma(\frac{1}{4})}.$

Type II:

Consider the definite integral

$$I_0 = \int_{-\infty}^{\infty} e^{-(x^4-2bx^2)}\,dx, \quad b>0$$

Integrate by parts letting $u = e^{-(x^4-2bx^2)}$ and $dv = dx$. Then

$$I_0 = \int_{-\infty}^{\infty} (4x^4 - 4bx^2)\, e^{-(x^4-2bx^2)}\, dx$$

$$= 4\int_{-\infty}^{\infty} x^4 e^{-(x^4-2bx^2)}\, dx - 4b\int_{-\infty}^{\infty} x^2 e^{-(x^4-2bx^2)}\, dx.$$

Now obviously $I_0$ cannot be zero. Hence dividing by $I_0$ we find $1 = 4\mu_4 - 4b\mu_2$ and therefore

$$b = \frac{4\mu_4 - 1}{4\mu_2} = \frac{\mu_4 - .25}{\mu_2}. \qquad (\mu_2 \text{ cannot be zero}).$$

Now that $b$ is known (calculated from the given data by this last formula) it is possible in any particular problem to find by mechanical quadrature the value of the integral $I_0$ to any desired degree of approximation. The simple rectangle formula with even a small number of ordinates known will give a good approximation.

Return now to the integration by parts just performed. The result takes the form

$$I_0 = \frac{d^2 I_0}{db^2} - 2b\frac{dI_0}{db} \qquad \text{or}$$

$$\frac{d^2 I_0}{db^2} - 2b\frac{dI_0}{db} - I_0 = 0$$

which is a Riccati[2] differential equation. Riccati's equation is

[2] Johnson's Differential Equations, p 227.

$\frac{d^2v}{dx^2}+2ax^{m-1}\frac{dv}{dx}+a(m-1)x^{m-2}v=0$. It has a solution capable of expression in finite form in terms of elementary functions if $m$ is the reciprocal of an odd positive integer. In our equation $m=2$, $a=-1$ hence no finite solution for the differential equation is possible. That is no finite expression in terms of elementary functions can be obtained for $I_o$. The solution of the Ricatti equation here is

$$I_o=C_1\left(1+\frac{b^2}{2!}+\frac{5\cdot1b^4}{4!}+\frac{9\cdot5\cdot1b^6}{6!}+\frac{13\cdot9\cdot5\cdot1b^8}{8!}+\dots\dots\right)$$

$$+C_2b\left(1+\frac{3b^2}{3!}+\frac{7\cdot3b^4}{5!}+\frac{11\cdot7\cdot3b^6}{7!}+\frac{15\cdot11\cdot7\cdot3b^8}{9!}+\dots\right). \tag{1}$$

To determine $C_1$ and $C_2$ we note that when $b=0$ then

$$I_o=\frac{1}{2}\Gamma\left(\frac{1}{4}\right)=C_1$$

and that when $b=0$ then $\left.\frac{dI_o}{db}\right]_{b=0}=2\int_{-\infty}^{\infty}x^2e^{-x^4}dx=\Gamma\left(\frac{3}{4}\right)=C_2$.

It is worth while to make certain transformations on the differential equation

$$\frac{d^2I_o}{db^2}-2b\frac{dI_o}{db}-I_o=0.$$

Let $I_o=e^{b^2/2}v$. Then

$$\frac{d^2v}{db^2}-b^2v=0.$$

Let $b^2/2=t$. Then

$$\frac{d^2v}{dt^2}+\frac{1}{2t}\frac{dv}{dt}-v=0.$$

Let $v=t^{1/4}w$. Then

$$t^2\frac{d^2w}{dt^2}+t\frac{dw}{dt}-\left[t^2+(1/4)^2\right]w=0.$$

Let $t=ix$ where $i=\sqrt{-1}$. Then

$$x^2\frac{d^2w}{dx^2}+x\frac{dw}{dx}+\left[x^2-(1/4)^2\right]w=0.$$

This last equation is Bessel's differential equation[8] with $n=1/4$. Hence its solution is

$$w=AJ_{\frac{1}{4}}(x)+BJ_{-\frac{1}{4}}(x)=AJ_{\frac{1}{4}}\left(\frac{-ib^2}{2}\right)+BJ_{-\frac{1}{4}}\left(\frac{-ib^2}{2}\right)\text{where}$$

$$J_n(x)=\frac{x^n}{2^n\Gamma(n+1)}\left(1+\frac{x^2}{(n+1)\cdot 2^2}+\frac{x^4}{(n+1)(n+2)\cdot 2^4\cdot 2!}+\cdots\right)$$

$$=\sum_{r=0}^{\infty}\frac{(-1)^r(x/2)^{n+2r}}{\Gamma(n+1+r)\cdot r!}.$$

The above transformations give

$$I_0=(b^2/2)^{1/4}e^{b^2/2}w.$$

Hence

$$I_0=(b^2/2)^{1/4}e^{b^2/2}\left[AJ_{\frac{1}{4}}\left(\frac{-ib^2}{2}\right)+BJ_{-\frac{1}{4}}\left(\frac{-ib^2}{2}\right)\right].$$

[8] Johnson's Differential Equations, p. 235.

Setting $b=0$ in $I_0$ we find $B=\dfrac{\frac{1}{2}\Gamma(\frac{1}{4})\Gamma(\frac{3}{4})}{\sqrt[4]{2i}}$.

Setting $b=0$ in $\dfrac{dI_0}{db}$ we find $A=\dfrac{2^{3/4}\Gamma(\frac{1}{4})\Gamma(\frac{3}{4})}{\sqrt[4]{-i}}$.

Putting in these values for $A$ and $B$ we find finally

$$I_0=e^{b^2/2}\left[\frac{1}{2}\Gamma(\tfrac{1}{4})\left(1+\frac{b^4}{3\cdot 4}+\frac{b^8}{7\cdot 3\cdot 4^2\cdot 2!}+\cdots\cdots\right)\right. \tag{2}$$

$$\left.+\Gamma(\tfrac{3}{4})b\left(1+\frac{b^4}{5\cdot 4}+\frac{b^8}{9\cdot 5\cdot 4^2\cdot 2!}+\cdots\cdots\right)\right].$$

It is worth noting, for purposes of computation, that the expression (2) converges much more rapidly than the form (1) given above,[4] on account of the factoring out of $e^{b^2/2}$. In addition the series in (2) have the advantage that the powers increase by 4 instead of by 2 as in (1). It will be shown presently that ordinarily $b$ is less than unity. But even for $b=1$ it will not be necessary to go further than the terms involving $b^{12}$ to get at least seven decimal places of accuracy. For $b$ less than one even fewer terms will suffice for this degree of accuracy.

[4] The form (1) is obtained immediately if we write

$$I_0=\int_{-\infty}^{\infty}e^{-(x^4-2bx^2)}dx=2\int_0^{\infty}e^{-(x^4-2bx^2)}dx=2\int_0^{\infty}e^{-x^4}e^{2bx^2}dx$$

$$=2\int_0^{\infty}e^{-x^4}\left(1+2bx^2+\frac{4b^2x^4}{2!}+\frac{8b^3x^6}{3!}+\cdots\cdots\right)dx,$$

assume term by term integration permissible, and make use of the fact

already mentioned that $\displaystyle\int_0^{\infty}x^pe^{-x^q}dx=\frac{1}{q}\Gamma\left(\frac{p+1}{q}\right)$.

From the point of view of the Ricatti differential equation it can be shown that $I_0 = \int_{-\infty}^{\infty} e^{-(x^4 - 2bx^2)} dx$ is the solution of

$$\frac{d^2 I_0}{db^2} - 2b\frac{dI_0}{db} - I_0 = 0$$

when the solution is sought in the form of a definite integral.[5] For the differential equation $b\phi(D)v + \psi(D)v = 0$ where $D = \frac{d}{db}$ and $\phi$ and $\psi$ are polynomials in $b$ with constant coefficients is satisfied by

$$v = c\int_{\alpha}^{\beta} e^{bt + \int \psi(t)T(t)dt} T(t)dt$$

where $c$ is a constant, $T(t)$ is the reciprocal of $\phi(t)$, and $\alpha$ and $\beta$ are so chosen that for all values of $b$

$$\left[e^{bt + \int \psi(t)T(t)dt}\right]_{\alpha}^{\beta} = 0.$$

Let $b\phi(D)v + \psi(D)v = D^2 v - 2bDv - v$. Then

$$\phi(t) = -2t,\ \psi(t) = t^2 - 1,$$

$$T(t) = -\frac{1}{2t},\ \int \psi(t)T(t)dt = -\frac{1}{2}\left(\frac{t^2}{2} - \log t\right),$$

and $\left[e^{bt - \frac{1}{2}\left(\frac{t^2}{2} - \log t\right)}\right]_{\alpha}^{\beta} = 0$ for all values of $b$ if $\alpha = 0, \beta = \infty$.

Hence

$$v = c\int_0^{\infty} e^{bt - \frac{1}{2}\left(\frac{t^2}{2} - \log t\right)}\left(-\frac{1}{2t}\right)dt$$

[5] A. R. Forsyth's Differential Equations, 6th edition, 1929, pp. 277-280.

$$= -\frac{c}{2}\int_0^\infty e^{-\frac{t^2}{4}-bt)}\frac{dt}{\sqrt{t}}.$$

Now let $t = 2x^2$ and $c = -\sqrt{2}$. Then

$$V = 2\int_0^\infty e^{-(x^4-2bx^2)}dx$$

$$= \int_{-\infty}^\infty e^{-(x^4-2bx^2)}dx$$

$$= I_0 .$$

An idea of the variation of $I_0$ as a function of $b$ can be obtained from the following table calculated from (1) for values of $b$ at intervals of 0.1 from 0 to 1 and using $\Gamma(\frac{1}{4}) = 3.625610$, $\Gamma(\frac{3}{4}) = 1.225417$. The results are plotted in the accompanying graph, Fig. 1.

| $b$ | $I_0$ |
|---|---|
| 0 0 | 1.812 805 |
| 0.1 | 1.945 063 |
| 0.2 | 2.099 726 |
| 0 3 | 2.282 225 |
| 0.4 | 2.499 648 |
| 0.5 | 2.761 349 |
| 0.6 | 3.079 783 |
| 0.7 | 3.471 748 |
| 0.8 | 3.960 152 |
| 0.9 | 4.576 578 |
| 1.0 | 5.365 158 |

The modes are at $x = \pm\sqrt{b}$ . Ordinarily the ordinate at the modes will not be greater than $e = 2.7182$ times the ordinate at $x = 0$ . Hence ordinarily it will not be necessary to consider values of $b$ greater than unity.

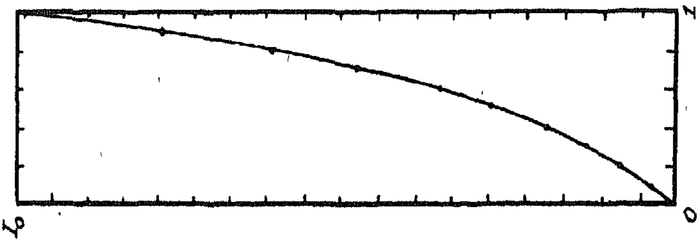



FIG. I

$$I_1 = \int_{-\infty}^{\infty} x e^{-(x^4-2bx^2)} dx = 0.$$

$$I_2 = \int_{-\infty}^{\infty} x^2 e^{-(x^4-2bx^2)} dx = \frac{1}{2}\frac{dI_0}{db}.$$

$$I_3 = \int_{-\infty}^{\infty} x^3 e^{-(x^4-2bx^2)} dx = 0.$$

$$I_4 = \int_{-\infty}^{\infty} x^4 e^{-(x^4-2bx^2)} dx = \left(\frac{1}{2}\right)^2 \frac{d^2 I_0}{db^2}.$$

$$I_{2n} = \int_{-\infty}^{\infty} x^{2n} e^{-(x^4-2bx^2)} dx = \left(\frac{1}{2}\right)^n \frac{d^n I_0}{db^n}, \; n = 0,1,2,3 \cdots$$

$$I_{2n+1} = \int_{-\infty}^{\infty} x^{2n+1} e^{-(x^4-2bx^2)} dx = 0, \; n = 0,1,2,3 \cdots$$

To find these derivatives one might use the relations[6]

$$\frac{dJ_n(x)}{dx} = \frac{n}{x} J_n(x) - J_{n+1}(x) = J_{n-1}(x) - \frac{n}{x} J_n(x).$$

But term by term differentiation is permissible and for this purpose it is simpler to use (1) rather than (2). We find

$$I_0 = \frac{1}{2}\Gamma\left(\frac{1}{4}\right)\left[1+\frac{b^2}{2!}+\frac{5b^4}{4!}+\frac{9\cdot5b^6}{6!}+\frac{13\cdot9\cdot5b^8}{8!}+\cdots\right]+\Gamma\left(\frac{3}{4}\right)\left[b+\frac{3b^3}{3!}+\frac{7\cdot3b^5}{5!}+\frac{11\cdot7\cdot3b^7}{7!}+\cdots\right],$$

$$\frac{dI_0}{db} = \frac{1}{2}\Gamma\left(\frac{1}{4}\right)\left[b+\frac{5b^3}{3!}+\frac{9\cdot5b^5}{5!}+\cdots\cdots\right]+\Gamma\left(\frac{3}{4}\right)\left[1+\frac{3b^2}{2!}+\frac{7\cdot3b^4}{4!}+\cdots\cdots\right],$$

$$\frac{d^2I_0}{db^2} = \frac{1}{2}\Gamma\left(\frac{1}{4}\right)\left[1+\frac{5b^2}{2!}+\frac{9\cdot5b^4}{4!}+\cdots\right]+\Gamma\left(\frac{3}{4}\right)\left[3b+\frac{7\cdot3b^3}{3!}+\frac{11\cdot7\cdot3b^5}{5!}+\cdots\right].$$

Since $b \geqq 0$ hence $I_0$ and all its derivatives are greater than zero.

Now the total probability is to be unity hence take $k = \frac{1}{I_0}$.

$$\mu_1 = \frac{I_1}{I_0} = 0,$$

$$\mu_2 = \frac{I_2}{I_0} = \frac{\frac{dI_0}{db}}{2I_0}.$$

$$\mu_3 = \frac{I_3}{I_0} = 0,$$

$$\mu_4 = \frac{I_4}{I_0} = \frac{\frac{d^2I_0}{db^2}}{4I_0},$$

etc.

Type III:

$$y = ke^{-(x^4 - 4cx)}.$$

[6] Whittaker and Watson, Modern Analysis, third edition, p. 360.

This curve is not symmetrical. But, obviously, changing $c$ to $-c$ has the same effect as changing $x$ to $-x$ or simply reversing the shape of the curve and the distribution from which it arises Hence it will be necessary to consider only positive values of $c$. As stated already, it is easy to show that there is a real mode at the point given by $x$ equal to the real cube root of $c$. If $c=1$ then $y=ke^{3}$, that is $e^{3}$ times the value of the ordinate at $x=0$. Hence usually $c$ will not be as great as unity.

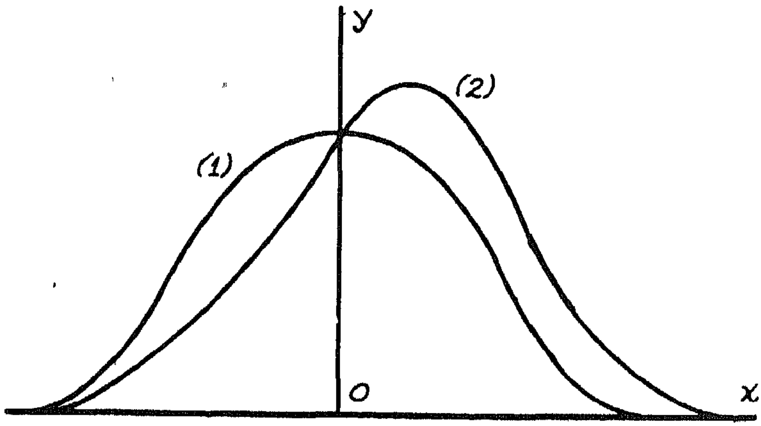



FIG. II

(1) $y=ke^{-x^4}$.  (2) $y=ke^{-(x^4-4cx)}$.

Let $I_0=\int_{-\infty}^{\infty} e^{-(x^4-4cx)}dx$.

Integrate by parts letting $u=e^{-x^4}$, $dv=e^{4cx}dx$.
Then

$$I_0=\frac{1}{c}\int_{-\infty}^{\infty} x^3 e^{-(x^4-4cx)}dx.$$

Hence

$$c = \frac{\int_{-\infty}^{\infty} x^3 e^{-(x^4 - 4cx)} dx}{I_0}$$

$$= \mu_3'.$$

With this value of $c$ calculated from the given data mechanical quadrature can be used to find an approximate value for $I_0$. Then let

$$k = \frac{1}{I_0}.$$

The result of the integration by parts could have been written in the form of the differential equation

$$\frac{d^3 I_0}{dc^3} - 64 c I_0 = 0.$$

Conversely, it is easy to show that $I_0 = \int_{-\infty}^{\infty} e^{-(x^4 - 4cx)} dx$ is the definite integral form of the solution of the differential equation

$$\frac{d^3 v}{dc^3} - 64 c v = 0.$$

For, here

$$\phi(D) = 64, \quad \psi(D) = D^3, \quad T(t) = -\frac{1}{64}, \quad \int \psi(t) T(t) dt = -\frac{t^4}{256},$$

and $\left[ e^{ct - t^4/256} \right]_{\alpha}^{\beta} = 0$ for all values of $c$ if $\alpha = -\infty$, $\beta = \infty$. Hence

$$V = -\frac{K}{64}\int_{-\infty}^{\infty} e^{ct - t^4/256}\,dt.$$

Now let $t = 4x$ and $c = -16$. Then

$$V = \int_{-\infty}^{\infty} e^{-(x^4 - 4cx)}\,dx$$

$$= I_0.$$

An expression for the value of $I_0$ can be obtained either by finding the series solution of the differential equation and determining the constants by setting $c = 0$ in $I_0$ and its derivatives, or by expanding $e^{4cx}$ in series in the definite integral itself and then integrating term by term.

$$I_0 = \int_{-\infty}^{\infty} e^{-(x^4 - 4cx)}\,dx$$

$$= \int_{-\infty}^{0} e^{-(x^4 - 4cx)}\,dx + \int_{0}^{\infty} e^{-(x^4 - 4cx)}\,dx$$

$$= \int_{0}^{\infty} e^{-(x^4 + 4cx)}\,dx + \int_{0}^{\infty} e^{-(x^4 - 4cx)}\,dx$$

$$= \int_{0}^{\infty} e^{-x^4}(e^{-4cx} + e^{4cx})\,dx = 2\int_{0}^{\infty} e^{-x^4}\cosh(4cx)\,dx$$

$$= \int_{0}^{\infty} e^{-x^4}\left[1 + \frac{(4cx)^2}{2!} + \frac{(4cx)^4}{4!} + \frac{(4cx)^6}{6!} + \cdots\cdots\right] dx$$

$$=2\sum_{n=0}^{\infty}\frac{(4c)^{2n}}{(2n)!}\int_0^{\infty}x^{2n}e^{-x^4}dx$$

$$=\frac{1}{2}\sum_{n=0}^{\infty}\frac{(4c)^{2n}}{(2n)!}\Gamma\frac{(2n+1)}{4}$$

$$=\frac{1}{2}\Gamma\left(\frac{1}{4}\right)\left[1+\frac{(4c)^4}{4\cdot 4!}+\frac{5(4c)^8}{4^2\cdot 8!}+\frac{9\cdot 5(4c)^{12}}{4^3\cdot 12!}+\ \cdots\ \right]$$

$$+\frac{1}{2}\Gamma\left(\frac{3}{4}\right)\left[\frac{(4c)^2}{2!}+\frac{3(4c)^6}{4\ 6!}+\frac{7\cdot 3(4c)^{10}}{4^2\cdot 10!}+\ \cdots\ \right].$$

These series may be differentiated term by term to obtain the derivatives of $I_o$ and hence

$$I_1=\int_{-\infty}^{\infty}xe^{-(x^4-4cx)}dx=\frac{1}{4}\frac{dI_o}{dc},$$

$$I_2=\int_{-\infty}^{\infty}x^2e^{-(x^4-4cx)}dx=\left(\frac{1}{4}\right)^2\frac{d^2I_o}{dc^2},$$

$$I_3=\int_{-\infty}^{\infty}x^3e^{-(x^4-4cx)}dx=\left(\frac{1}{4}\right)^3\frac{d^3I_o}{dc^3},$$

$$I_4=\int_{-\infty}^{\infty}x^4e^{-(x^4-4cx)}dx=\left(\frac{1}{4}\right)^4\frac{d^4I_o}{dc^4},$$

If $x$ is replaced by $x-\sqrt[3]{c}$ the effect is to translate the modal value of $x$ to the origin. The equation of the curve then becomes

$$y=\frac{1}{I_0}e^{-(x^4+c_1x^3+c_2x^2+c_3x+c_4)}$$

where

$$c_1=-4\sqrt[3]{c},$$

$$c_2=6\sqrt[3]{c^2},$$

$$c_3=-8c,$$

$$c_4=5c\sqrt[3]{c}.$$

Type IV:

$$y=ke^{-(x^4+px^2+qx)}$$

Consider the definite integral

$$I_0=\int_{-\infty}^{\infty}e^{-(x^4+px^2+qx)}dx$$

If $p=q=0$ we get Type I. If $p\neq 0, q=0$ we get Type II. If $p=0$, $q\neq 0$ we get Type III. Hence consider now $p\neq 0,\ q\neq 0$.

Integrate $I_0$ by parts with $u=e^{-(x^4+px^2)}$ and $dv=e^{-qx}dx$. Then

$$I_0=-\frac{1}{q}\int_{-\infty}^{\infty}(4x^3+2px)e^{-(x^4+px^2+qx)}dx$$

$$=-\frac{4}{q}\int_{-\infty}^{\infty}x^3e^{-(x^4+px^2+qx)}dx-\frac{2p}{q}\int_{-\infty}^{\infty}xe^{-(x^4+px^2+qx)}dx.$$

Now divide by $I_0$ and multiply by $q$. Then

$$q = -4\mu_3' - 2p\mu_1'.$$

Begin again with $I_0$ and integrate by parts, this time with

$u = e^{-(x^4+px^2+qx)}$ and $dv = dx$. Then

$$I_0 = \int_{-\infty}^{\infty} (4x^4 + 2px^2 + qx) e^{-(x^4+px^2+qx)} dx$$

$$= 4\int_{-\infty}^{\infty} x^4 e^{-(x^4+px^2+qx)} dx + 2p\int_{-\infty}^{\infty} x^2 e^{-(x^4+px^2+qx)} dx$$

$$+ q\int_{-\infty}^{\infty} x e^{-(x^4+px^2+qx)} dx.$$

Divide by $I_0$. Then

$$1 = 4\mu_4' + 2p\mu_2' + q\mu_1'.$$

Now substitute $q = -4\mu_3' - 2p\mu_1'$ in $1 = 4\mu_4' + 2p\mu_2' + q\mu_1'$ and we get

$$p = \frac{1 + 4\mu_1'\mu_3' - 4\mu_4'}{2(\mu_2' - \mu_1'^2)},$$

(3)

$$q = -4\mu_3' - 2p\mu_1' = -4\mu_3' - \mu_1'\left(\frac{1 + 4\mu_1'\mu_3' - 4\mu_4'}{\mu_2' - \mu_1'^2}\right).$$

The result of the two integrations by parts can be written in the form of two simultaneous partial differential equations. They are

$$4\frac{\partial^2 I_o}{\partial p\,\partial q} - 2p\frac{\partial I_o}{\partial q} + qI_o = 0$$

$$4\frac{\partial^2 I_o}{\partial p^2} - 2p\frac{\partial I_o}{\partial p} - q\frac{\partial I_o}{\partial q} - I_o = 0.$$

Let $S_o = \int_{-\infty}^{\infty} e^{-(x^4+px^2)}\,dx.$ Then

$$S_{2n} = \int_{-\infty}^{\infty} x^{2n} e^{-(x^4+px^2)}\,dx = (-1)^n\frac{d^n S_o}{dp^n},$$

$$S_{2n+1} = \int_{-\infty}^{\infty} x^{2n+1} e^{-(x^4+px^2)}\,dx = 0.$$

$$I_o = \int_{-\infty}^{\infty} e^{-(x^4+px^2+qx)}\,dx$$

$$= \int_{-\infty}^{\infty} e^{-(x^4+px^2)} e^{-qx}\,dx$$

$$= \int_{-\infty}^{\infty} e^{-(x^2+px^2)}\left[1-(qx)+\frac{(qx)^2}{2!}-\frac{(qx)^3}{3!}+\cdots\right]dx$$

$$= \sum_{i=0}^{\infty} \frac{q^{2i}}{(2i)!} S_{2i} .$$

$$I_{2n} = \int_{-\infty}^{\infty} x^{2n} e^{-(x^4+px^2+qx)} dx$$

$$= (-1)^n \frac{\partial^n I_0}{\partial p^n}$$

$$= \sum_{i=0}^{\infty} \frac{q^{2i}}{(2i)!} S_{2n+2i}$$

$$I_{2n+1} = \int_{-\infty}^{\infty} x^{2n+1} e^{-(x^4+px^2+qx)} dx$$

$$= (-1) \frac{\partial I_{2n}}{\partial q}$$

$$= \sum_{i=0}^{\infty} (-1) \frac{q^{2i-1}}{(2i-1)!} S_{2n+2i}$$

When the values of $p$ and $q$ are calculated from the data of any given problem by the formulas (3) then values for $I_0$, $I_1$, $I_2$, $I_3$, $I_4$ . etc. can be obtained by mechanical quadrature.

For two real, distinct modes $-8p^3>27q^2, (p<0)$. Hence if $-2<p<0$ then $1.54>q>1.54$. If $-8p^3=27q^2$ then one mode flattens forming a point of inflexion with a horizontal tangent at the minimum point. Changing $q$ to $-q$ has the same effect as changing $x$ to $-x$ and hence $q$ is a component of skewness of the curve. If the curve is so placed that the sum of the values of $x$ at the

modes and at the minimum point is zero then the equation of the curve will be of the form

$$y = ke^{-(x^4+px^2+qx)}$$

If now we change the scale of $x$ by replacing $x$ by $x\sqrt{a}$ then we are led to the functions of the form

$$y = ke^{-a^2(x^4+px^2+qx)}.$$

Performing the two integrations by parts, as before, on the integral

$$I = \int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)}\,dx$$

leads to the relations

$$\begin{aligned} q &= -4\mu_3' - 2p\mu_1' = -4(\mu_3 + 3M\mu_2 + M^3) - 2Mp, \\ p &= \frac{\frac{1}{a^2} + 4\mu_1'\mu_3' - 4\mu_4'}{2(\mu_2' - \mu_1'^2)} = \frac{\frac{1}{a^2} - 4(\mu_4 + 3M\mu_3 + 3M^2\mu_2)}{2\mu_2}. \end{aligned} \tag{4}$$

If

$$\int_{-\infty}^{\infty} x^n e^{-(x^4+px^2+qx)}\,dx = I_n'(p, q), \quad n = 0, 1, 2, 3 \ldots.$$

then

$$\int_{-\infty}^{\infty} x^n e^{-a^2(x^4+px^2+qx)}\,dx = \frac{1}{a^{(n+1)/2}} I_n'(ap, a^{3/2}q) = I_n(ap, a^{3/2}q).$$

In particular,

$$I_0 = I_0(0, 0) = \int_{-\infty}^{\infty} e^{-a^2x^4}\,dx = \frac{1}{2\sqrt{a}}\Gamma\left(\frac{1}{4}\right),$$

$$I_{2n-1} = I_{2n-1}(0,0) = 0,$$

$$I_{2n} = I_{2n}(0,0) = \int_{-\infty}^{\infty} x^{2n} e^{-a^2x^4} dx = \frac{1}{2a^{(2n+1)/2}} \Gamma\left(\frac{2n+1}{4}\right).$$

$$\mu_1 = I_1 / I_0 = 0,$$

$$\mu_2 = I_2 / I_0 = \frac{1}{a} \frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)},$$

$$\mu_3 = I_3 / I_0 = 0,$$

$$\mu_4 = I_4 / I_0 = \frac{1}{4a^2}$$

$$\mu_{2n-1} = I_{2n-1} / I_0 = 0,$$

$$\mu_{2n} = I_{2n} / I_0 = \frac{1}{a^n} \frac{\Gamma\left(\frac{2n+1}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}.$$

In the case of Type I when

$$y = y_0 e^{-a^2x^4}, \quad y_0 = \frac{1}{2\sqrt{a}} \Gamma\left(\frac{1}{4}\right),$$

$p = q = 0$ and hence from (4), or as can be shown directly, $a^2 = \frac{1}{4\mu_4}$.

In Type II,

$$y = y_0 e^{-a^2(x^4 + px^2)}, \quad \frac{1}{y_0} = \int_{-\infty}^{\infty} e^{-a^2(x^4+px^2)} dx,$$

$q=0$ and hence $a^2 = \dfrac{1}{2p\mu_2 + 4\mu_4}$ . In the Type III where

$y = y_0 e^{-a^2(x^4+qx)}, \dfrac{1}{y_0} = \int_{-\infty}^{\infty} e^{-a^2(x^4+qx)} dx$, $p=0$ and hence

$$a^2 = \frac{1}{4(\mu_4 + 3M\mu_3 + 3M^2\mu_2)}.$$

In general, since $p$ and $q$ are determined when the modes and minimum point of the curve are known, theoretically at least, $a^2$ is fixed by the relations (4). In practice, however, this would mean that the accuracy in the determination of $a^2$ would be contingent upon the accuracy with which the modes and minimum point are determined. Hence other methods for fixing $a^2$ will be required in general. Now if in $I_0(ap, a^{3/2}q)$ we replace $p$ and $q$ by (4) which involve only $a^2$ and quantities calculable from the given data we have a function of $a$ alone, say $f(a)$. It will be sufficient then if we determine a value of $a$ such that $f(a) = N$ where $N$ is the total given frequency Then fix $p$ and $q$ by (4) and the modes and minimum point by $4x^3 + 2px + q = 0$.

The points of inflexion are found from the equation

$$\frac{d^2y}{dx^2} = 0$$

and for Type I are given by $x^4 = \dfrac{3}{4a^2}$ . Hence

$$x = \frac{\pm 0.930605}{\sqrt{a}}$$

approximately. For Type II they are given by $8a^2x^6 + 8a^2px^4 + 2(a^2p^2-3)x^2 - p = 0$. For Type III they are given by $16a^2x^6 + 8a^2qx^3 - 12x^2 + a^2q^2 = 0$. And in general they are given by roots of the equation

$$16a^2x^6 + 16a^2px^4 + 8a^2qx^3 + 4(a^2p^2-3)x^2 + 4a^2pqx + a^2q^2 - 2p = 0.$$

It will be noticed that the distribution given by

$$y = y_0 e^{-a^2(x^4 + px^2 + qx)}$$

can have the Mean at the origin if and only if $q = O$, that is, if and only if the distribution is symmetrical. Now replace $x$ by $x - m$. The area remains the same and hence also $y_0$.
The equation then is

$$y = y_0 e^{-a^2(x^4 + p_1 x^3 + p_2 x^2 + p_3 x + p_4)} \text{ where}$$

$$p_1 = -4m,$$

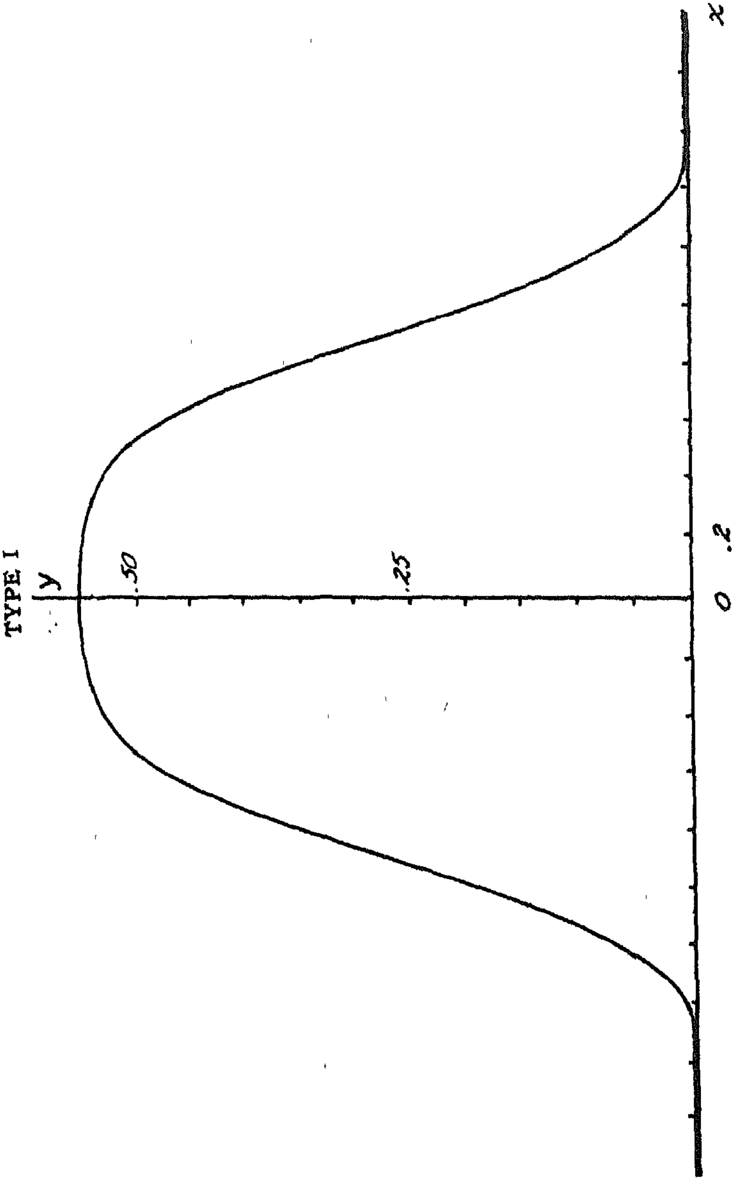
$$p_2 = 6m^2 + p,$$

$$p_3 = q - 2mp - 4m^3,$$

$$p_4 = m^4 + m^2 p - mq,$$

and $p$ and $q$ are given by the relations (4) above. An integration by parts with $u = e^{-a^2(x^4 + p_1 x^3 + p_2 x^2 + p_3 x + p_4)}$ shows that

$$a^2(4\mu_4' + 3p_1\mu_3' + 2p_2\mu_2' + p_3\mu_1') = 1.$$


TYPE I
y
.50
.25
0
.2
x

This small beginning of the study of the system of frequency curves with the quartic exponent will be concluded here with the construction of artificial illustrations of Types I and II.

TYPE I

$$y = \frac{2}{\Gamma(\frac{1}{4})} e^{-x^4}, \quad \log_{10} \Gamma(\tfrac{1}{4}) = 0.5593811,$$

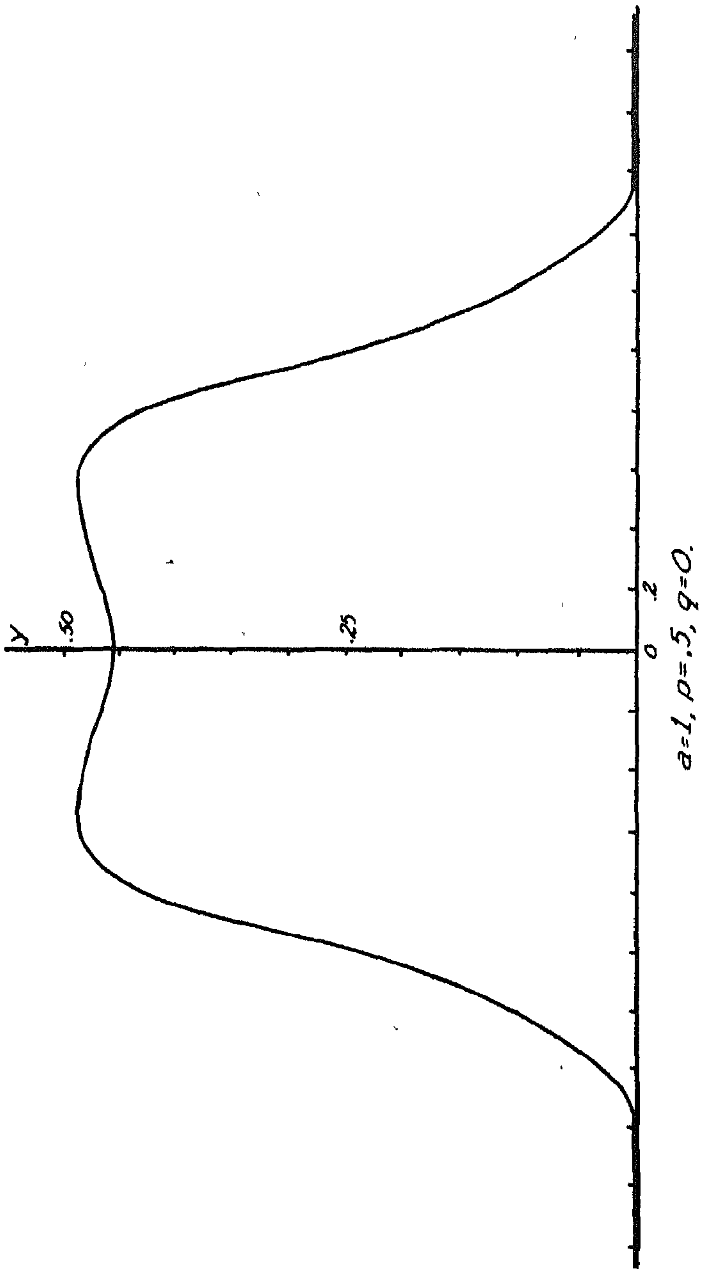
| $x$ | $y$ | $x^2y$ | $x^4y$ |
|---|---|---|---|
| 0.0 | 0.5516313 | 0.0000000 | 0.0000000 |
| 0.1 | .5515762 | .0055158 | .0000551 |
| 0.2 | .5507494 | 0220300 | .0008812 |
| 0.3 | 5471811 | .0492463 | .0044322 |
| 0 4 | 5376888 | 0860302 | .0137648 |
| 0.5 | .5182096 | .1295524 | .0323881 |
| 0.6 | .4845787 | .1744483 | .0628014 |
| 0 7 | .4338852 | .2126037 | .1041758 |
| 0.8 | .3662367 | .2343915 | 1500105 |
| 0.9 | .2862255 | .2318426 | .1877925 |
| 1.0 | .2029338 | .2029338 | .2029338 |
| 1.1 | .1275846 | .1543774 | .1867966 |
| 1.2 | .0693579 | .0998754 | .1438205 |
| 1.3 | .0317147 | .0535978 | .0905803 |
| 1.4 | .0118376 | .0232017 | .0454753 |
| 1 5 | .0034917 | .0078563 | .0176767 |
| 1.6 | .0007861 | .0020124 | .0051518 |
| 1.7 | .0001301 | .0003760 | .0010866 |
| 1 8 | 0000152 | .0000492 | 0001596 |
| 1.9 | .0000012 | 0000043 | .0000156 |
| 2.0 | .0000001 | .0000004 | .0000016 |
| | 5.2758155 | 1.6899455 | 1.2500000 |

Total frequency $= \frac{2(5.2758155) - 0.5516313}{10} = 1.0000000$

$$\mu_2 = \frac{2(1.6899455) - 0.0000000}{10} = 0.3379891$$

$$\mu_4 = \frac{2(1.2500000) - 0.0000000}{10} = 0.2500000.$$

TYPE II



$$y = \frac{1}{2.187099} e^{-(x^4 - 0.5x^2)}.$$

TYPE II

| $x$ | $y$ | $x^2y$ | $x^4y$ |
|---|---|---|---|
| 0.0 | 0.4572267 | 0.0000000 | 0.0000000 |
| 0.1 | .4594725 | .0045947 | .0000459 |
| 0.2 | .4657175 | .0186287 | .0007451 |
| 0.3 | .4744135 | .0426972 | .0038427 |
| 0.4 | 4827888 | .0772462 | .0123594 |
| 0.5 | .4867153 | .1216788 | 0304197 |
| 0.6 | .4808614 | .1731101 | .0623196 |
| 0.7 | .4594725 | 2251415 | .1103193 |
| 0.8 | .4180410 | .2675462 | .1712296 |
| 0.9 | .3556970 | .2881146 | .2333728 |
| 1.0 | .2773220 | .2773220 | .2773220 |
| 1 1 | .1936552 | .2343228 | 2835306 |
| 1.2 | .1181056 | .1700721 | 2449038 |
| 1.3 | 0611958 | 1034209 | 1747813 |
| 1.4 | .0261429 | .0512401 | .1004306 |
| 1.5 | .0089145 | .0200576 | .0451297 |
| 1.6 | .0023433 | .0059988 | .0153571 |
| 1.7 | .0004575 | 0013222 | .0038211 |
| 1.8 | .0000638 | .0002067 | 0006697 |
| 1.9 | .0000061 | .0000220 | .0000795 |
| 2.0 | .0000004 | .0000016 | .0000064 |
| | 5.2286133 | 2 0827448 | 1 7706859 |

Total frequency $= \dfrac{2(5.2286133) - 0.4572267}{10} = 1.000000,$

$$\mu_2 = \frac{2(2.0827448) - 0.0000000}{10} = 0.41654896$$

$$\mu_4 = \frac{2(1.7706859) - 0.0000000}{10} = 0.35413718.$$

From relations (1) or (2) it is found that when $b = 0.25$, (i.e. $p = -0.5, q = 0$)

then $I_0 = 2.187099$; Conversely, the formula $b = \dfrac{\mu_4 - 0.25}{\mu_2}$ gives, retaining six decimal places, $b = 0.250000.$

(To be Continued in May Issue)

# ON THE LOGARITHMIC FREQUENCY DISTRIBUTION AND THE SEMI-LOGARITHMIC CORRELATION SURFACE*

*By*

PAE-TSI YUAN

## INTRODUCTION**

The method of treating frequency curves as developed chiefly by Edgeworth, Kapteyn, Van Uven and Wicksell occupies an important place in both theoretical and applied statistics. The essence of this method may be briefly summarized as follows.

Suppose a function of the variable $z$ is distributed according to the normal law of error. Then, $z$ certainly cannot be also normally distributed, unless the function is a linear function of $z$. Without losing generality, we shall write the normally distributed function in standard units as $x = f(z)$. Thus the origin of $x$ is its mean and the unit of $x$ is its standard deviation. The relative frequency of values of $x$ between $x$ and $x + dx$ is, therefore

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad .$$

and the relative frequency of values of $z$ between $z$ and $z+dz$ is

$$\frac{1}{\sqrt{2\pi}} f'(z) e^{-\frac{1}{2}\left[f(z)\right]^2} dz .$$

Thus if we have an observed frequency distribution of $z$ and we know a normally distributed function of $z$, then we can

* A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the University of Michigan.

** Papers written by the writers mentioned in this introduction are listed under the writers' names in the Bibliography at the end of this paper.

graduate the distribution of $z$ by using this formula. Edgeworth calls this method of graduating a frequency distribution the method of translation. In two papers on "Skew Frequency Curves in Biology and Statistics" published in 1903 and 1916, J. C. Kapteyn elegantly set forth a theoretical foundation of this method. Later Wicksell gave a similar justification. Both of them based their "genetic theory of frequency", to use Wicksell's terminology, upon a generalized hypothesis of elementary errors.

In the present paper, we are interested only in the important special case where $x = \frac{1}{c} \log \frac{z-a}{b}$ The frequency function of $z$, then, becomes:

$$\frac{1}{\sqrt{2\pi}\, c(z-a)}\, e^{-\frac{1}{2c^2}\left(\log \frac{z-a}{b}\right)^2}$$

which is called the logarithmic frequency function.*

Numerous papers have been written on this frequency curve. Among the early writers were Francis Galton and McAllister. But a systematic treatment on the properties of this curve from the standpoint of mathematical statistics is still lacking. Hence, in the first part of this paper, such a treatment will be given, thus leading to some interesting relationships among the characteristics of this curve.

Various methods of determining the parameters of this frequency function have been proposed by writers on this subject. Pearson is the first writer to make use of the method of moments. Later this method was also applied by Jørgensen and Wicksell. In this paper, the method of moments will be considered and a table will be provided to facilitate the computation of the constants by this method.

Edgeworth, Wicksell and Van Uven all have contributed in

---

* For a justification of this frequency function based on Weber-Fechner's Psychophysical Law see the "Calculus of Observations" by E. T. Whittaker and G. Robinson, pp. 217-218. (Blackie & Son Ltd., London and Glasgow, 1929)

extending the method of translation to correlation surfaces. Wicksell's logarithmic correlation surface is particularly noteworthy. In the last part of this paper, a semi-logarithmic correlation surface of two variables will be developed and its properties studied.

The writer wishes to express his appreciation for the assistance Professor Cecil C. Craig has given him in making this study.

## PART I

### THE LOGARITHMIC FREQUENCY DISTRIBUTION

For the sake of clarity, it is desirable to state at the outset that the logarithmic frequency distribution represented by

$$F(x) = \frac{1}{\sqrt{2\pi}\,c(x-a)}\, e^{-\frac{1}{2c^2}\left(\log \frac{x-a}{b}\right)^2} \tag{1}$$

is unimodal and has three parameters. The parameter $a$ is the finite lower or upper limit of $x$ according to whether $b$ is positive or negative. In the following discussions, unless the sign of $b$ plays an important rôle, we shall take $b$ to be positive and $a$ to be the finite lower limit of $x$. However, the results of our discussions can be easily modified to cover the case where $b$ is negative and $a$ is the finite upper limit of $x$.

In the first eight sections of Part I the properties of the logarithmic frequency distribution will be treated from the standpoint of mathematical statistics,* and in section 9 the numerical application of this distribution will be discussed.

### 1. *AVERAGES*

We shall first give the analytic expressions of four different averages of $x$ and then observe their relative magnitudes.

* Some topics under consideration here in regard to the properties of the logarithmic frequency distribution have also been discussed by many writers, among whom we may particularly mention McAllister, Kapteyn, Pearson and Pretorius. See the references under these writers names in the Bibliography of this paper.

By definition, the arithmetic mean of $z$ is

$$m = \int_a^\infty z F(z)\,dz = b e^{\frac{c^2}{2}} + a.$$

The logarithm of the geometric mean of $z$ about the point $z = a$ is given by

$$\int_a^\infty \log(z-a) F(z)\,dz = \log b.$$

Hence, the geometric mean of $z$ about $z = a$ measured from $z = 0$ is

$$m_g = b + a.$$

Since the median of $z$ corresponds to $x = \frac{1}{c}\log\frac{z-a}{b} = 0$, it is equal to

$$m_d = b + a.$$

Setting the derivative $\dfrac{dF(z)}{dz} = -\dfrac{\log\frac{z-a}{b} + c^2}{c^2(z-a)} F(z)$

equal to zero, we obtain the mode of $z$ as

$$m_o = b e^{-c^2} + a.$$

Thus, the geometric mean and the median are equal. Moreover,

$$m_o < m_d = m_g < m$$

2. *POINTS OF INFLECTION*

The second derivative of $F(z)$ is

$$\frac{d^2F(z)}{dz^2} = \frac{\left(\log\frac{z-a}{b}\right)^2 + 3c^2 + \log\frac{z-a}{b} - c}{c^4(z-a)^2} F(z).$$

The roots of the equation

$$\left(\log \frac{x-a}{b}\right)^2 + 3c^2 \log \frac{x-a}{b} - c^2 = 0$$

are the points of inflection of the logarithmic frequency curve. We shall denote them by $\breve{x}_1$ and $\breve{x}_2$,

$$\breve{x}_1 = b e^{-\frac{3}{2}c^2\left[1+\sqrt{1+\frac{4}{9c^2}}\right]} + a$$

$$\breve{x}_2 = b e^{-\frac{3}{2}c^2\left[1-\sqrt{1+\frac{4}{9c^2}}\right]} + a$$

Note that the quantity under the radical sign is always positive and greater than one. Its square root is, therefore, greater than one in absolute value. Hence, $\breve{x}_1 < b+a < \breve{x}_2$ . That is, the geometric mean and the median of $x$ lie between the points of inflection.

Furthermore, if we observe that the points of inflection may be written in relation to the mode as

$$\breve{x}_1 - a = (m_0 - a) e^{-\frac{c^2}{2}\left(1+3\sqrt{1+\frac{4}{9c^2}}\right)}$$

$$\breve{x}_2 - a = (m_0 - a) e^{-\frac{c^2}{2}\left(1-3\sqrt{1+\frac{4}{9c^2}}\right)}$$

we see that $\breve{x}_1 < m_0 < x_2$ .

But the mean does not always lie between the two inflection points, since

$$\breve{x}_1 - a = (m-a) e^{-\frac{c^2}{2}\left(4+3\sqrt{1+\frac{4}{9c^2}}\right)}$$

$$\breve{x}_2 - a = (m-a) e^{-\frac{c^2}{2}\left(4-3\sqrt{1+\frac{4}{9c^2}}\right)} .$$

Obviously, $\breve{z}_1$ is always less than the mean. But when $c^2 > 4/7$, the mean is situated above both points of inflection.

Now, the relation of the averages and the points of inflection, when $c^2 < 4/7$, may be expressed by the inequality

$$\breve{z}_1 < m_o < m_d = m_g < m < \breve{z}_2$$

which holds for almost all practical cases, since $c^2$ rarely exceeds $4/7$ in practice.

3. *HIGH CONTACT*

A frequency function is said to have high contact, if the function and all its derivatives vanish at the upper and the lower limits of the variable $z$. We know that the logarithmic frequency function vanishes at both the finite and the infinite limits of $z$. It can be easily seen that all its derivatives also vanish at these points, if we make the substitution $z' = \log \frac{z-a}{b}$, which will throw every derivative of the logarithmic frequency function into a product of two factors, one being a polynomial in $z'$ and another being $e^{-\frac{1}{2c^2}z'^2 + kz'}$ where $k$ is a positive integer. Thus, it is obvious that all the derivatives become zero, as $z'$ approaches $\pm\infty$, which correspond to the finite and the infinite limits of $z$. For instance, this substitution will put the first derivative of the logarithmic frequency function $F(z)$,

$$\frac{dF(z)}{dz} = -\frac{\log \frac{z-a}{c} + c^2}{c^2(z-a)} F(z)$$

into the form

$$\frac{z' - c^2}{\sqrt{2\pi}\, c^3 b^2}\, e^{-\frac{1}{2c^2}z'^2 + 2z'}$$

which clearly goes to zero as $z'$ approaches $\pm\infty$, that is, as $z$

approaches "$a$" and infinity.

The logarithmic frequency function, therefore, has high contact.

## 4. MOMENTS

We shall study the practical application of the method of moments to determine the parameters of the logarithmic frequency distribution in section 9. But at present we must know the relationships between the parameters and the moments in order to discuss the properties of dispersion, skewness and kurtosis

First, we shall express the moments in terms of the parameters·

The $s$-$th$ moment of $z$ about the point $z=a$ is given by

$$\mu_s' = \int_a^{\infty} (z-a)^s F(z)\,dz = b^s e^{\frac{s^2c^2}{2}}.$$

And we also have the recurring relation $\mu_s' = b e^{\frac{(2s-1)c^2}{2}} \mu_{s-1}'$.

The $s$-$th$ moment of $z$ about the mean is

$$\mu_s = \int_a^{\infty} (z-m)^s F(z)\,dz = b^s e^{\frac{sc^2}{2}} \sum_{k=0}^{s} (-1)^{s-k} \binom{s}{k} e^{\frac{k(k-1)c^2}{2}}.$$

Consequently, the $s$-$th$ standard moment of $z$, $\alpha_s = \frac{\mu_s}{\mu_2^{s/2}}$ is

$$\alpha_s = (e^{c^2}-1)^{-\frac{s}{2}} \sum_{k=0}^{s} (-1)^{s-k} \binom{s}{k} e^{\frac{k(k-1)c^2}{2}}.$$

Setting $s$ equal to 3 and 4, we have

$$\alpha_3 = \pm (e^{c^2}-1)^{\frac{1}{2}} (e^{c^2}+2)$$

$$\alpha_4 = 3 + (e^{c^2}-1)(e^{3c^2} + 3e^{2c^2} + 6e^{c^2} + 6)$$

which will be discussed in connection with skewness and kurtosis. Note that the sign of $\alpha_3$ follows that of $b$, because the sign of the third moment of $x$ about the mean is determined by $b^3$.

Now, we want to express the parameters in terms of the moments. It is clear that there is an infinite number of ways to accomplish this, since there is an infinitude of moments. But we are particularly interested in the expressions of the parameters in terms of the mean and the second and third moments about the mean. Letting $\omega = e^{c^2}$, we have

$$
\begin{aligned}
m &= b\omega^{\frac{1}{2}} + a \\
\mu_2 &= b^2\omega(\omega - 1) \\
\mu_3 &= b^3\omega^{3/2}(\omega-1)^2(\omega+2).
\end{aligned}
\tag{2}
$$

Solving these equations for the parameters, we find $\omega$ is the only real root of the cubic.

$$
\omega^3 + 3\omega^2 - (4 + \alpha_3^2) = 0 \tag{3}
$$

Hence, the parameters $c$, $b$ and $a$ may be expressed as

$$
\begin{aligned}
c &= (\log \omega)^{\frac{1}{2}} \\
b &= \left(\frac{1}{\omega-1}\right)^{\frac{1}{2}}\left(\frac{1}{\omega}\right)^{\frac{1}{2}}\sigma = \left(\frac{\omega+2}{\alpha_3}\right)\left(\frac{1}{\omega}\right)^{\frac{1}{2}}\sigma \\
a &= m - \left(\frac{1}{\omega-1}\right)^{\frac{1}{2}} = m - \left(\frac{\omega+2}{\alpha_3}\right)\sigma
\end{aligned}
\tag{4}
$$

where the sign of $b$ follows that of $\alpha_3$, and $\sigma = \sqrt{\mu_2}$. The practical application of (3) and (4) will be discussed in section 9. We shall now turn our attention to some other properties of the logarithmic frequency distribution.

## 5. *DISPERSION*

The dispersion of $x$ about the mean may be measured by the standard deviation, $\sigma = \sqrt{\mu_2} = be^{\frac{c^2}{2}}(e^{c^2}-1)^{\frac{1}{2}}$. Denote the

deviation of $z$ from the mean in terms of the standard deviation by $t=(z-m)/\sigma$. Then, with the aid of formulae (1), (2) and (4), we obtain the distribution of $t$ as

$$\frac{(e^{c^2}-1)^{\frac{1}{2}}}{\sqrt{2\pi}\,c\left[1+(e^{c^2}-1)^{\frac{1}{2}}t\right]}\, e^{-\frac{1}{2c^2}\left\{\log\left[1+(e^{c^2}-1)^{\frac{1}{2}}t\right]+\frac{c^2}{2}\right\}^2} dt \qquad (5)$$

where $(e^{c^2}-1)^{\frac{1}{2}}=\alpha_3/(e^{c^2}+2)$ takes the same sign as $\alpha_3$.

We know that for the normal distribution 50% of the total frequency lies between the limits $t=-.6745$ and $t=+.6745$. Now, we want to know the similar limits of $t$ for the logarithmic distribution. For that reason, we write $t$ directly in terms of the normally distributed function $x=\frac{1}{c}\log\frac{z-a}{b}$

$$t=\frac{z-m}{\sigma}=\frac{(e^{xc-\frac{c^2}{2}}-1)}{(e^{c^2}-1)^{\frac{1}{2}}}. \qquad (6)$$

Placing $x$ equal to $-.6745$ and $+.6745$ we have at once the limits

$$t_1=\frac{(e^{-.6745c-\frac{c^2}{2}}-1)}{(e^{c^2}-1)^{\frac{1}{2}}}$$

$$t_2=\frac{(e^{.6745c-\frac{c^2}{2}}-1)}{(e^{c^2}-1)^{\frac{1}{2}}}$$

between which 50% of the total frequency is included. These limits are two quartiles and obviously depend on $c$. It is clear that one can also locate other deciles and percentiles of $t$ by using (6).

An abstract measure of the dispersion is the coefficient of variability which expresses the standard deviation in terms of the

mean. For the logarithmic distribution, it is

$$D=\left|\frac{\sigma}{m-a}\right|=\left|(e^{c^2}-1)^{\frac{1}{2}}\right| \tag{7}$$

which shows that in a logarithmic distribution the larger $c^2$ is, the greater is the variability.

It is interesting to note that if we also express the deviation of $x$ from the mean in terms of $(m-a)$ and denote it by

$$t'=\frac{x-m}{m-a}=(e^{c^2}-1)^{\frac{1}{2}}t$$

we have by (5) the distribution of $t'$ in this simple form

$$\frac{1}{\sqrt{2\pi}c(t'+1)}e^{-\frac{1}{2c^2}\left[\log(1+t')+\frac{c^2}{2}\right]^2}dt'. \tag{8}$$

6 SKEWNESS

It has been proposed to use $\alpha_3/2$ or $\alpha_3$ as a measure of skewness of a frequency distribution. For the logarithmic curve, we have shown that

$$\alpha_3=(e^{c^2}-1)^{\frac{1}{2}}(e^{c^2}+2)$$

or

$$\alpha_3=(\omega-1)^{\frac{1}{2}}(\omega+2). \tag{9}$$

Hence, the absolute value of $\alpha_3$ increases with $c$. Since $c$ can take on any finite value whatever, the skewness of the logarithmic curve as measured by $\alpha_3$ can also have any finite value. Moreover, as we have seen, $\alpha_3$ of the logarithmic distribution can be positive as well as negative.

In Figure 1 are shown four logarithmic curves with $m=0$, $\sigma=1$ and with varying $\alpha_3$'s Various parameters calculated from formulae (4) and important characteristics of these curves are exhibited in Table I.

When $c=0$, $\alpha_3$ also vanishes. In fact, the logarithmic curve approaches the normal curve of error, as $c$ goes to zero. This can be demonstrated as follows With the aid of formulae (4) we can write the normally distributed function $x=\frac{1}{c}\log\frac{z-a}{b}$ as

$$x=\frac{1}{c}\log\frac{z-a}{b}$$

$$=\frac{1}{c}\left\{\frac{c^2}{2}+\log\left[1+\frac{z-m}{\sigma}(e^{c^2}-1)^{\frac{1}{2}}\right]\right\}$$

$$=\frac{c}{2}\;\frac{z-m}{\sigma}\;\frac{(e^{c^2}-1)^{\frac{1}{2}}}{c}-\frac{(z-m)^2}{2\sigma^2}\;\frac{(e^{c^2}-1)}{c}+\cdots$$

Now, it can be easily seen that

$$\lim_{c\to 0} x = \frac{z-m}{\sigma}$$

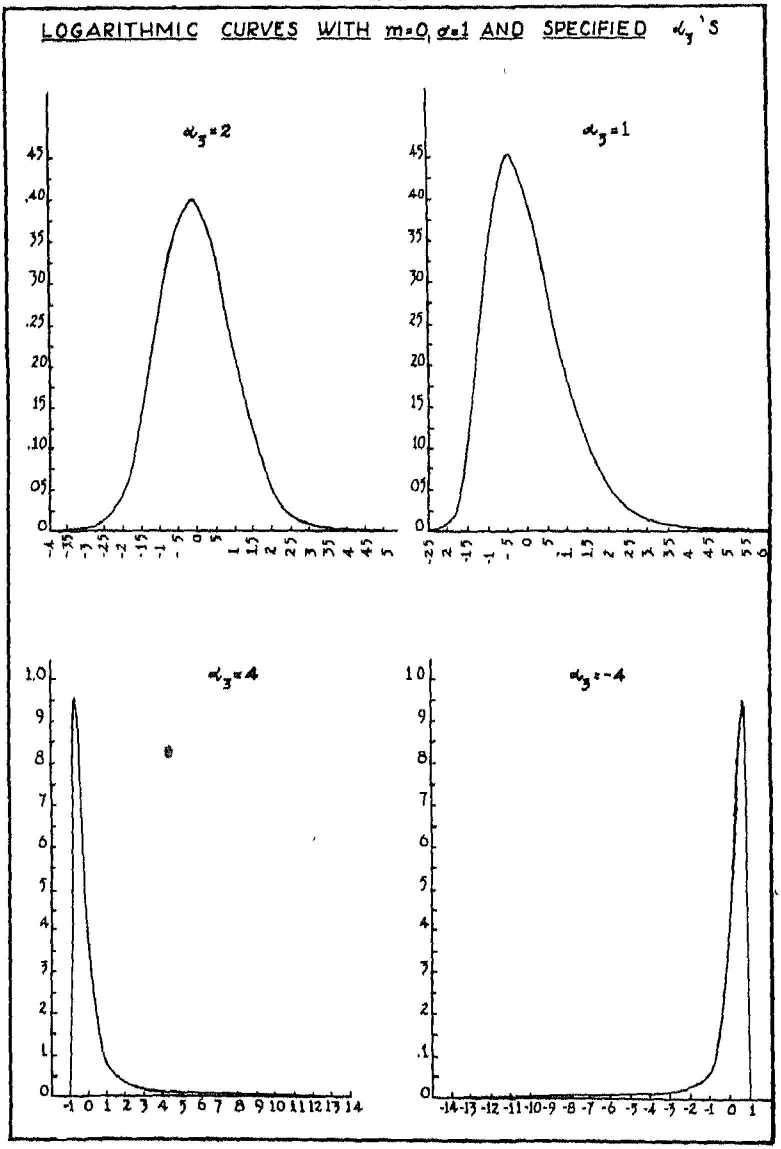
which is a linear function of $z$. Hence, the logarithmic distribution of $z$ approaches the normal distribution as $c$ approaches zero.

TABLE I

Parameters and Important Characteristics of the Logarithmic Curves with $m=0$, $\sigma=1$ and Specified $\alpha_3'$s

| $\alpha_3$ | 2 | 1 | 4 | - 4 |
|---|---|---|---|---|
| $\omega$ | 1.0044 | 1.1038 | 2.0000 | 2.0000 |
| $c$ | .0665 | .3143 | .8326 | .8326 |
| $a$ | -15.0222 | -3.1038 | -1.0000 | 1.0000 |
| $b$ | 14.9890 | 2.9543 | 7071 | - .7071 |
| $m_g=m_d$ | - .0332 | -.1495 | - .2929 | .2929 |
| $m_o$ | - .0991 | - .4274 | - .6465 | .6465 |
| $\check{z}_1$ | - 1.09 | -1 30 | - 9341 | .9341 |
| $\check{z}_2$ | .90 | .50 | - .0532 | .0532 |
| $D$ | .0666 | .3221 | 1.0000 | 1.0000 |
| $\alpha_4-3$ | .0712 | 1.8295 | 35.0000 | 35.0000 |

FIGURE I


LOGARITHMIC CURVES WITH m=0, σ=1 AND SPECIFIED α₃'S
$\alpha_3=2$
$\alpha_3=1$
$\alpha_3=4$
$\alpha_3=-4$

Another measure of skewness is defined by Pearson as

$$\chi = \frac{m - m_0}{\sigma} .$$

For the logarithmic curve, it becomes

$$\chi = \frac{1 - \omega^{-\frac{3}{2}}}{(\omega - 1)^{\frac{1}{2}}}$$

which has a maximum value equal to $.6561$, when $\omega = 1.7202$ and $\alpha_3 = 3.1573$ This, however, does not indicate that the skewness of the logarithmic curve is limited. Rather it shows that $\chi$ is not a satisfactory measure of skewness, so far as the logarithmic curve is concerned For any measure of skewness should characterize the skewness of a curve without ambiguity, and $\chi$ fails to do so in case of the logarithmic curve. For instance, when we say that a certain logarithmic curve has $\chi = .32$, we may mean either a logarithmic curve with $\alpha_3 = .68$ or one with $\alpha_3 = 36.00$.

When the logarithmic curve is only moderately skew, $\chi$ approximately equals $\alpha_3/2$. This can be shown as follows: Letting $h^2 = \omega - 1$, we have

$$\chi = \frac{1 - (1 + h^2)^{-\frac{3}{2}}}{h} = \frac{3}{2}h - \frac{15}{8}h^3 + \frac{35}{16}h^5 \dots$$

and $\alpha_3 = 3h + h^3$.

Hence, for small $|h|$ and hence small $|\alpha_3|$ $\chi$ approximately equals $\alpha_3/2$. For instance, when $\alpha_3 = .2$, $\chi = .0991$ which is approximately $\alpha_3/2 = .1$

We may mention here that for the Pearsonian type III curve, the relation $\chi = \alpha_3/2$ always holds. In fact, it appears from Table II that the type III curve and the logarithmic curve are very similar for small $|\alpha_3|$. But the differences between them are already pronounced for $\alpha_3 = 1$, as we can see from Table III.

TABLE II

Ordinates and Areas of the Logarithmic Curve and the Pearsonian Type III Curve

$m = 0 \quad \sigma = 1 \quad \alpha_3 = .2$

| $x$ | Ordinate at $x$ | | Area from the Lower Limit to $x$ | |
|---|---|---|---|---|
| | Log Curve | Type III | Log Curve | Type III |
| -3.5 | .0003 | .0002 | 0000 | .0000 |
| -3.0 | 0020 | 0020 | 0005 | 0004 |
| -2.5 | 0124 | 0123 | 0034 | 0034 |
| -2.0 | .0491 | 0492 | 0172 | 0171 |
| -1.5 | .1337 | .1341 | 0607 | .0607 |
| -1.0 | 2587 | 2591 | .1579 | 1582 |
| -.5 | .3692 | 3687 | .3178 | 3172 |
| 0 | .3991 | 3986 | 5132 | .5133 |
| .5 | .3366 | .3364 | .7006 | .7002 |
| 1.0 | .2267 | 2267 | .8418 | .8417 |
| 1.5 | 1242 | 1245 | .9285 | 9284 |
| 2.0 | .0568 | 0567 | .9720 | .9721 |
| 2.5 | 0217 | .0217 | .9906 | 9906 |
| 3.0 | 0072 | 0071 | 9972 | 9973 |
| 3.5 | 0020 | .0020 | 9993 | .9993 |
| 4.0 | 0006 | .0005 | .9998 | .9998 |
| 4.5 | 0002 | .0001 | | |

TABLE III

Ordinates and Areas of the Logarithmic Curve and the Pearsonian Type III Curve

$m = 0 \qquad \sigma = 1 \qquad \alpha_3 = 1$

| $x$ | Ordinate at $x$ | | Area from the Lower Limit to $x$ | |
|---|---|---|---|---|
| | Log. Curve | Type III | Log. Curve | Type III |
| - 2.0 | .0084 | 0 | .0009 | 0 |
| - 1.5 | .1196 | .1226 | .0259 | .0190 |
| - 1.0 | .3364 | .3609 | .1398 | .1429 |
| - .5 | .4498 | .4481 | .3442 | .3528 |
| 0 | .4040 | .3907 | .5624 | .5665 |
| .5 | .2883 | .2807 | .7363 | .7345 |
| 1.0 | .1791 | .1785 | .8520 | .8488 |
| 1.5 | .1017 | .1043 | .9210 | .9182 |
| 2.0 | .0548 | .0573 | .9590 | .9576 |
| 2.5 | .0295 | .0300 | .9783 | .9788 |
| 3.0 | .0144 | .0151 | .9895 | .9897 |
| 3.5 | .0073 | .0074 | .9948 | .9951 |
| 4.0 | .0036 | .0035 | .9977 | .9977 |
| 4.5 | .0017 | .0017 | .9987 | .9990 |
| 5.0 | .0009 | .0008 | .9993 | .9995 |
| 5.5 | .0004 | .0003 | .9997 | .9998 |
| 6.0 | .0002 | .0002 | .9998 | .9999 |
| 6.5 | .0001 | .0001 | .9999 | |

## 7. KURTOSIS

Another important characteristic of a frequency curve is kurtosis measured by $\frac{1}{8}(\alpha_4-3)$ or simply by $\eta=\alpha_4-3$, which equals zero for the normal law of error. If the mean and the standard deviation are taken to be the origin and the unit, respectively, then usually the frequency of a curve in the vicinity of the mean is in excess or in defect to that of a normal curve according to whether $\eta$ is positive or negative. A curve is said to be platykurtic, if $\eta>0$. It is leptokurtic, if $\eta<0$. Thus, the logarithmic curve is always platykurtic, for its $\eta$ is

$$\eta=(\omega-1)(\omega^3+3\omega^2+6\omega+6) \tag{10}$$

or

$$\eta=\omega^4+2\omega^3+3\omega^2-6$$

and $\omega>1$ Since the logarithmic curve has only three parameters, there exists a functional relationship between its skewness and kurtosis. This relationship is given through the parameter $\omega$ by (9) and (10). We may further deduce the following relations from these two equations:

$\eta$ is always greater than $\frac{3}{2}\alpha_3^2$. This follows from the fact that $2\eta-3\alpha_3^2=(\omega-1)(2\omega^3+3\omega^2)>0$.

For $|\alpha_3|<6.44$, we have $3\alpha_3^2>\eta$, since $3\alpha_3^2-\eta=(\omega-1)(-\omega^3+6\omega+6)>0$ holds, provided $\omega<2.8$, which corresponds to $|\alpha_3|<6.44$.

For $|\alpha_3|<2.15$, we have $2\alpha_3^2>\eta$, since $2\alpha_3^2-\eta=(\omega-1)(-\omega^3-\omega^2+2\omega+2)>0$ holds provided $\omega<1.4$, which corresponds to $|\alpha_3|<2.15$.

Since practically the value of $|\alpha_3|$ can hardly reach $6.44$ or even $2.15$, the relations just stated hold for all practical instances.

The relationship existing between $\eta$ and $\alpha_3$ is sometimes used as a criterion for applying the logarithmic curve to observed data We shall discuss this point in section 9

## 8. POWERS, ROOTS AND PRODUCTS OF THE LOGARITHMICALLY DISTRIBUTED VARIABLES

If $z$ is logarithmically distributed and has "$a$" as its lower limit, $W=(z-a)^k$ is also so distributed, $k$ being any constant. This follows from the fact that if $x=\frac{1}{c}\log\frac{z-a}{b}$ is normally distributed, so is $kx=\frac{k}{c}\log\frac{z-a}{b}$. From the frequency function of $z$, $F(z)$, given by (1), we find at once the analytic expression of the frequency distribution of $W$ to be

$$\frac{1}{\sqrt{2\pi}\,ckW}\,e^{-\frac{1}{2c^2k^2}\left[\log\frac{W}{b^k}\right]^2}dW. \tag{11}$$

We have learned from the preceding sections that a logarithmic distribution represented by (1) with larger $c$ has greater variability, skewness, and kurtosis. Thus, if $k^2>1$, the variability, skewness, and kurtosis are greater for $W$ than for $z$. On the other hand, if $k^2<1$, the distribution of $z$ has greater variability, skewness, and kurtosis.

If the logarithmically distributed variables $z_1, z_2, \cdots, z_n$ are independent and have for their lower limits, $a_1, a_2, \cdots, a_n$, then the product

$$Y=(z_1-a_1)(z_2-a_2)\cdots(z_n-a_n)$$

is also so distributed. This follows from the fact that if

$$x_1=\frac{1}{c_1}\log_1\frac{z_1-a_1}{b_1},\quad x_2=\frac{1}{c_2}\log\frac{z_2-a_2}{b_2},\ldots,x_n=\frac{1}{c_n}\log\frac{z_n-a_n}{b_n}$$

are each normally distributed and are independent, their sum also obeys the normal law of error.

Since the variables are independent, the frequency distribu-

tion of these $n$ variables is represented by

$$F_1(z_1)F_2(z_2)\cdots F_n(z_n)\,dz_1\,dz_2\cdots dz_n \tag{12}$$

where $F_i(z_i)=\dfrac{1}{\sqrt{2\pi}\,c_i(z_i-a_i)}\,e^{-\frac{1}{2c_i^2}\left[\log\frac{z_i-a_i}{b_i}\right]^2}$.

Substituting $z_1-a_1=Y/(z_2-a_2)\cdots(z_n-a_n)$ in (12) and integrating the resulting expression with respect to $z_2,\cdots,z_n$ successively over the respective ranges, we have the distribution of $Y$ as

$$\frac{1}{\sqrt{2\pi}\sqrt{c_1^2+c_2^2+\cdots+c_n^2}\,Y}\,e^{-\frac{1}{2(c_1^2+c_2^2+\cdots+c_n^2)}\left[\log\frac{Y}{b_1b_2\cdots b_n}\right]^2}dY. \tag{13}$$

Since the sum, $c_1^2+c_2^2+\cdots+c_n^2$, is greater than any $c_i^2$, the distribution of $Y$ has greater variability, skewness, and kurtosis than that of each individual variable.

## 9. *NUMERICAL APPLICATIONS*

Many methods of fitting a logarithmic frequency curve to observed data have been proposed. But only the method of moments will be considered below.*

The method of moments is very simple to apply. It consists of placing the computed moments in equations (2) and then determining the parameters by solving these equations by formulae (3) and (4).† The only step of computation which requires some time and care to obtain accurate results is the solution of

* Among other methods of graduating the logarithmic frequency distribution, the graphical method proposed by Kapteyn and Van Uven is especially useful. For a description of this method, refer to their paper on "Skew Frequency Curves in Biology and Statistics, 2nd Paper".

† In his paper, "On the Genetic Theory of Frequency", Wicksell also showed the application of the method of moments to the logarithmic frequency distribution. However, he found the parameter "a" first and then proceeded to obtain "log b" and "c".

the cubic.

$$\psi(\omega)=\omega^3+2\omega^2-(\alpha_3^2+4)=0$$

Hence, it is desirable to have a table which will provide an approximation of the required root of this cubic for a given $\alpha_3$. Then, the root can be approximated to as great a degree of accuracy as we wish by applying, for instance, Newton's method. That is why Table IV is constructed. Practically, after we obtain an approximate value of $\omega$ from Table IV, one single application of Newton's method will almost invariably suffice to give us a value of $\omega$ accurate to four decimal places. In Table IV, values of $c$ corresponding to given values of $\omega$ are also provided to serve as a check to our computation of $c$ by formulae (4).

TABLE IV

Table Facilitating the Solution of the Cubic

$$\omega^3+3\omega^2-(\alpha_3^2+4)=0$$

| $\omega$ | $\alpha_3$ | $c$ | $\omega$ | $\alpha_3$ | $c$ |
|---|---|---|---|---|---|
| 1. | 0 | 0 | | | |
| 1.01 | .3010 | .1000 | 1.26 | 1.6623 | .4807 |
| 1.02 | .4271 | .1407 | 1.27 | 1.6991 | .4889 |
| 1.03 | .5248 | .1720 | 1.28 | 1.7356 | .4969 |
| 1.04 | .6080 | .1980 | 1.29 | 1.7717 | .5046 |
| 1.05 | .6820 | .2209 | 1.30 | 1.8075− | .5122 |
| 1.06 | .7495+ | .2415− | 1.31 | 1.8429 | .5196 |
| 1.07 | .8122 | .2602 | 1.32 | 1.8781 | .5269 |
| 1.08 | .8712 | .2775− | 1.33 | 1.9129 | .5340 |
| 1.09 | .9270 | .2936 | 1.34 | 1.9475+ | .5410 |
| 1.10 | .9803 | .3087 | 1.35 | 1.9819 | .5478 |
| 1.11 | 1.0315− | .3231 | 1.36 | 2.0160 | .5545+ |
| 1.12 | 1.0808 | .3366 | 1.37 | 2.0499 | .5611 |
| 1.13 | 1.1285+ | .3496 | 1.38 | 2.0836 | .5675+ |
| 1.14 | 1.1749 | .3619 | 1.39 | 2.1171 | .5738 |
| 1.15 | 1.2200 | .3739 | 1.40 | 2.1503 | .5801 |
| 1.16 | 1.2640 | .3852 | 1.41 | 2.1835− | .5862 |
| 1.17 | 1.3070 | .3962 | 1.42 | 2.2164 | .5922 |
| 1.18 | 1.3492 | .4068 | 1.43 | 2.2492 | .5981 |
| 1.19 | 1.3905− | .4171 | 1.44 | 2.2818 | .6038 |
| 1.20 | 1.4311 | .4270 | 1.45 | 2.3143 | .6096 |
| 1.21 | 1.4710 | .4366 | 1.46 | 2.3467 | .6151 |
| 1.22 | 1.5103 | .4460 | 1.47 | 2.3789 | .6207 |
| 1.23 | 1.5491 | .4550 | 1.48 | 2.4110 | .6261 |
| 1.24 | 1.587 | .4638 | 1.49 | 2.4430 | .6315+ |
| 1.25 | 1.6250 | .4723 | 1.50 | 2.4749 | .6368 |

To illustrate the use of Table IV and to help in studying the application of the logarithmic frequency curve, we take the distribution of the weights of 1,000 female students from the "Synopsis of Elementary Mathematical Statistics"* by Miss B. L. Shook. (See Table V.)

The mean, standard deviation, and skewness for this distribution† are

$$m = 118.74 \text{ lbs.}$$

$$\sigma = 16\ 91752 \text{ lbs.}$$

$$\alpha_3 = .976424$$

To compute $\omega$, we find from Table IV that for $\alpha_3 = .976424$ $\omega$ is approximately $\omega_0 = 1.10$. For a better approximation, we apply Newton's method ·

$$\omega = \omega_0 - \frac{\psi(\omega_0)}{\psi'(\omega_0)} = \omega_0 - \frac{\omega_0^3 + 3\omega_0^2 - (\alpha_3^2 + 4)}{3\omega_0^2 + 6\omega_0}$$

$$= 1.10 - \frac{.007596}{10.23} = 1.10 - .000743$$

$$= 1.099257$$

By formulae (4), the parameters $c$, $b$ and $a$ are found to be

$$c = .307627$$

$$b = 51.2160 \text{ lbs.}$$

$$a = 65.0423 \text{ lbs.}$$

* Annals of Mathematical Statistics, Vol. I, No. 1 (1930), p. 39.

† Sheppard's corrections have been duly applied.

TABLE V

Observed and Theoretical Distributions of the Weights of 1,000 Female Students

(Original Measurements Made to Nearest 1/10 lb.)

| Class Limits (Pounds) | Observed Frequency | Theoretical Logarithmic Distribution By Areas | Theoretical Logarithmic Distribution By Ordinates | Theoretical Type III Distribution By Areas |
|---|---|---|---|---|
| 70- 79.9 | 2 | 0 | 0 | 0 |
| 80- 89.9 | 16 | 10 | 6 | 4 |
| 90- 99.9 | 82 | 97 | 94 | 102 |
| 100-109.9 | 231 | 228 | 234 | 238 |
| 110-119.9 | 248 | 255 | 259 | 250 |
| 120-129.9 | 196 | 190 | 190 | 184 |
| 130-139 9 | 122 | 114 | 111 | 111 |
| 140-149.9 | 63 | 57 | 57 | 59 |
| 150-159.9 | 23 | 27 | 27 | 29 |
| 160-169.9 | 5 | 12 | 12 | 13 |
| 170-179.9 | 7 | 6 | 6 | 6 |
| 180-189.9 | 1 | 2 | 2 | 3 |
| 190-199.9 | 2 | 1 | 1 | 1 |
| 200-209.9 | 1 | 1 | 1 | 0 |
| 210-219-9 | 1 | 0 | 0 | 0 |
| Total | 1,000 | 1,000 | 1,000 | 1,000 |

Knowing $c$ , $b$ and $a$ , we obtain the geometric mean and the mode:

$$m_g = m_d = 116.2583 \text{ lbs.}$$
$$m_o = 111.6326 \text{ lbs.}$$

Using these parameters, the theoretical distribution of the weights of 1,000 female students has been computed and is shown in Table V and Figure II. The fit of the logarithmic distribution to the observed data is, indeed, excellent.* The lowest possible weight of female students, according to the theoretical distribution, is 65.04 pounds, which is just about what one would expect after examining the observed data.

Miss Shook† used the type III distribution to fit the same set of observed data, and gave the result as shown in the last col-

* Grouping the first three classes into one class and the last six classes into one class, we apply the $\chi^2$ test for goodness of fit and find that the probability to get a worse fit is .70.

† Annals of Mathematical Statistics, Vol. I, No. 3 (1930), p 242.

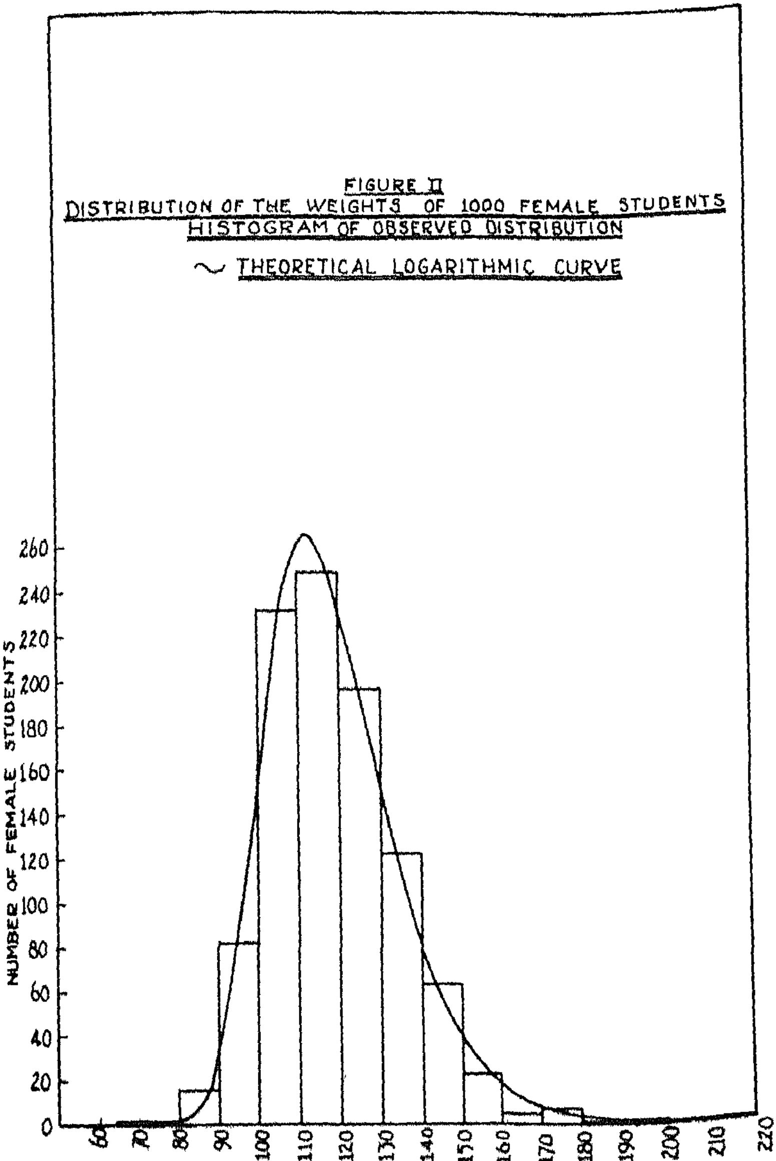


FIGURE II
DISTRIBUTION OF THE WEIGHTS OF 1000 FEMALE STUDENTS
HISTOGRAM OF OBSERVED DISTRIBUTION
THEORETICAL LOGARITHMIC CURVE
NUMBER OF FEMALE STUDENTS
0
20
40
60
80
100
120
140
160
180
200
220
240
260
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
POUNDS

umn of Table V. The fit is not as good as that given by the logarithmic distribution, especially in view of the fact that the type III curve fixes the least possible weight at 84.09 pounds, while as a matter of fact there are two students whose weights are below that limit.‡

From the standpoint of the method of moments, a criterion for the logarithmic distribution to fit a set of observed data is that $\eta = \alpha_4 - 3$ computed directly from the observed data must be approximately the same as the theoretical $\eta$ computed from formula (10). This criterion, however, does not seem to work in practice. For instance, for the distribution of the weights of 1,000 female students, the theoretical $\eta$ is 1.7419, while the observed $\eta$ is 2.4536. But in spite of this fact, the observed distribution, as we have seen, is very satisfactorily fitted by a logarithmic distribution.

Another criterion is to require the observed moments about the lower limit "$a$" to satisfy approximately the recurring relation

$$\mu_s' = be^{\frac{2s-1}{2}c^2}\mu_{s-1}'$$

for $s = 4$. This criterion is approximately fulfilled by the distribution of the weights of 1,000 female students, for which we have

$$\mu_4' = 147279 \cdot 10^2$$

$$be^{\frac{7}{2}c^2}\mu_3' = 146696 \cdot 10^2$$

and

$$\mu_4' / be^{\frac{7}{2}c^2}\mu_3' = 1.0040.$$

The fact that a set of observed data may be satisfactorily graduated by the logarithmic distribution but fulfills only the second criterion may be explained on the ground that the com-

‡ In fact, since the finite limit of the variable for type III curve is $m - \frac{2}{\alpha_3}\sigma$ and for the logarithmic curves $m - \frac{2+\omega}{\alpha_3}\sigma$, the finite limit is always greater in absolute value for the logarithmic curve than for the type III curve

paratively wide discrepancy between the observed and theoretical frequency in the classes near the lower limit maks a great difference in the fourth moment about the mean but does not make much difference in the fourth moment about the point "$a$".

## PART II

### THE SEMI-LOGARITHMIC CORRELATION SURFACE

Suppose that the correlation surface of the functions, $x = f(u,v)$ and $y = g(u,v)$, is a normal correlation surface and each has its mean as the origin and its standard deviation as the unit. Then, the probability that values of $x$ will lie between $x$ and $x + dx$ and values of $y$ between $y$ and $y + dy$ is

$$\phi(x,y)dxdy = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[x^2 - 2rxy + y^2\right]} dxdy. \quad (1)$$

It follows that the probability that values of $u$ will lie between $u$ and $u+du$ and values of $v$ between $v$ and $v+dv$ is $F(u,v)\,dudv$ given by

$$\frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[f^2 - 2rfg + g^2\right]} \begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix} du\,dv. \quad (2)$$

$F(u,v)$ is, therefore, a generalized correlation surface of two variables, deduced by extending the method of translation for treating frequency distributions of one variable.

It is clear that in this general form the correlation surface represented by $F(u,v)$ is of little practical use, on account of its complexity. Now a natural simplification suggests itself. That is to take $x$ as a function of $u$ only and $y$ as a function of $v$

only. By virtue of this simplification, $F(u, v)$ becomes

$$F(u,v)=\frac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left[f^2-2rfg+g^2\right]}\frac{df}{du}\frac{dg}{dv} \tag{3}$$

which is a great deal easier to handle than before.

Professor Wicksell has made use of (3) for the special case where, in our notations, $x$ and $y$ are

$$x=\frac{1}{c_1}\log\frac{u-a_1}{b_1}$$

$$y=\frac{1}{c_2}\log\frac{v-a_2}{b_2}$$

which leads to the so-called "logarithmic correlation surface" * The surface possesses the property that its marginal distributions as well as the distributions of $u$ for given values of $v$ and distributions of $v$ for given values of $u$ are all logarithmic frequency distributions.

Presently we shall study another case for which

$$x=\frac{u-\gamma}{\lambda}$$

$$y=\frac{1}{c}\log\frac{v-a}{b}.$$

The correlation surface $F(u,v)$ given by (3) then becomes:

$$F(u,v)=\frac{e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{u-\gamma}{\lambda}\right)^2-2r\frac{u-\gamma}{\lambda c}\log\frac{v-a}{b}+\left(\frac{1}{c}\log\frac{v-a}{b}\right)^2\right]}}{2\pi\lambda c(v-a)\sqrt{1-r^2}} \tag{4}$$

* In Wicksell's paper, "On the Genetic Theory of Frequency", the theory of the logarithmic correlation function is developed. In his two successive papers quoted in the Bibliography of this paper, the original theory is extended and the application of the extended results illustrated.

which may be appropriately called a semi-logarithmic correlation surface. We shall investigate its marginal distributions, moments and regression curves of the characteristics.

## 1. *MARGINAL DISTRIBUTIONS*

Now, we shall first find the distribution of the marginal totals of $u$. This can be, of course, accomplished very easily by integrating $F(u,v)$ with respect to $v$ over the range from $a$ to infinity. The result is:

$$\int_a^\infty F(u,v)dv = \frac{1}{\sqrt{2\pi}\lambda} e^{-\frac{1}{2\lambda^2}(u-\gamma)^2} \tag{5}$$

Thus, the marginal distribution of $u$ obeys the normal laws of error.

Similarly, if we integrate $F(u,v)$ with respect to $u$ over the range from $-\infty$ to $\infty$, we find at once the marginal distribution of $v$ as follows:

$$\int_{-\infty}^\infty F(u,v)du = \frac{1}{\sqrt{2\pi}\,c(v-a)} e^{-\frac{1}{2c^2}\left(\log\frac{v-a}{b}\right)^2} \tag{6}$$

which is, clearly a logarithmic distribution and, therefore, has all the properties and characteristics discussed in Part I. Hence, the semi-logarithmic correlation surface is characterized by the fact that one marginal distribution is normal, while the other is logarithmic. It is needless to mention that this does not constitute a sufficient condition for a correlation surface to be a semi-logarithmic correlation surface defined by (4).

## 2. *MOMENTS*

The moment, $\mu'_{ij}$, of the semi-logarithmic correlation surface about the point $u=\gamma$ and $v=a$ is given by

$$\begin{aligned}\mu'_{ij} &= \int_{-\infty}^\infty\int_a^\infty (u-\gamma)^i(v-a)^j F(u,v)\,du\,dv \\ &= \lambda^i b^j e^{\frac{j^2c^2}{2}} \sum_{k=0}^{i} \binom{i}{k} (jcr)^{i-k} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} t^k e^{-\frac{t^2}{2}} dt\end{aligned} \tag{7}$$

where $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} t^k e^{-\frac{t^2}{2}} dt = \frac{k!}{2^{k/2}\left(\frac{k}{2}\right)!}$ if $k$ is even

$= 0,$ if $k$ is odd.

Using relation (7), we can easily calculate the following six moments about the mean of $u$, $m_u$, and the mean of $v$, $m_v$:

$$\begin{aligned}
\mu_{10} &= m_u - \gamma = 0 \\
\mu_{20} &= \lambda^2 \\
\mu_{01} &= m_v - (be^{\frac{c^2}{2}} + a) = 0 \\
\mu_{02} &= b^2 e^{c^2}(e^{c^2} - 1) \\
\mu_{03} &= b^3 e^{\frac{3}{2}c^2}(e^{c^2} - 1)^2(e^{c^2} + 2) \\
\mu_{11} &= r\lambda c b e^{\frac{c^2}{2}} \quad .
\end{aligned} \tag{8}$$

Now, we want to solve these equations for the six parameters. As before, we let $\omega = e^{c^2}$ and write $\alpha_{03} = \mu_{03}/\mu_{02}^{\frac{3}{2}}$. Again, we have $\omega$ as the only real root of the cubic.

$$\omega^3 + 3^2 - (\alpha_{03}^2 + 4) = 0 \quad . \tag{9}$$

The six parameters of the semi-logarithmic correlation surface can be written as:

$$\begin{aligned}
\gamma &= m_u \\
\lambda &= \sqrt{\mu_{20}} = \sigma_u \\
c &= (\log \omega)^{\frac{1}{2}} \\
b &= \left(\frac{1}{\omega}\right)^{\frac{1}{2}}\left(\frac{1}{\omega - 1}\right)^{\frac{1}{2}}\sigma_v = \left(\frac{1}{\omega}\right)^{\frac{1}{2}}\left(\frac{\omega + 2}{\alpha_{03}}\right)\sigma_v \\
a &= m_v - \left(\frac{1}{\omega - 1}\right)^{\frac{1}{2}}\sigma_v = m_v - \left(\frac{\omega + 2}{\alpha_{03}}\right)\sigma_v \\
r &= \frac{\mu_{11}}{\sigma_u \sigma_v} \frac{(\omega - 1)^{\frac{1}{2}}}{(\log \omega)^{\frac{1}{2}}}
\end{aligned} \tag{10}$$

which furnish us a simple practical method for determining the parameters of the semi-logarithmic correlation surface for observed data.

## 3. *REGRESSION OF THE MEAN*

First, let us observe that the function $F(u,v)$ may be put into the following forms:

$$F(u,v)=\frac{1}{2\pi\sqrt{1-r^2}\,c(v-a)\lambda}\,e^{-\frac{1}{2c^2}\left(\log\frac{v-a}{b}\right)^2}\,e^{-\frac{1}{2(1-r^2)}\left[\frac{u-\gamma}{\lambda}-\frac{r}{c}\log\frac{v-a}{b}\right]^2}$$

$$=\frac{1}{2\pi\sqrt{1-r^2}\,c(v-a)\lambda}\,e^{-\frac{1}{2\lambda^2}(u-\gamma)^2}\,e^{-\frac{1}{2(1-r^2)}\left[\frac{1}{c}\log\frac{v-a}{b}-r\frac{u-\gamma}{\lambda}\right]^2}$$

Hence, the distribution of $u$ for a particular array of $v$ is normal.

$$\theta_1(u,v)=\frac{1}{\sqrt{2\pi}\,\lambda\sqrt{1-r^2}}\,e^{-\frac{1}{2(1-r^2)}\left[\frac{u-\gamma}{\lambda}-\frac{r}{c}\log\frac{v-a}{b}\right]^2} \tag{11}$$

and the distribution of $v$ for a particular array of $u$ is logarithmic:

$$\theta_2(u,v)=\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}\,c(v-a)}\,e^{-\frac{1}{2(1-r^2)}\left[\frac{1}{c}\log\frac{v-a}{b}-r\frac{u-\gamma}{\lambda}\right]^2}$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}\,c(v-a)}\,e^{-\frac{1}{2c^2(1-r^2)}\left[\log\frac{v-a}{be^{cr\frac{u-\gamma}{\lambda}}}\right]^2}. \tag{12}$$

To find the mean of $u$ for a particular value of $v$, we multiply $\theta_1(u,v)$ by $u$ and integrate the resulting expression with respect to $u$ over the range from $-\infty$ to $\infty$.

$$\bar{u}=\int_{-\infty}^{\infty}u\,\theta_1(u,v)\,du=\frac{\lambda r}{c}\log\frac{v-a}{b}+\gamma \tag{13}$$

which is the regression equation of the mean of $u$ on $v$ and may be called the logarithmic regression equation.

Similarly, the regression equation of the mean of $v$ on $u$ is found to be:

$$\bar{v} = \int_a^\infty v\theta_2(u,v)dv = be^{cr\frac{u-\gamma}{\lambda} + \frac{c^2}{2}(1-r^2)} + a \qquad (14)$$

which may be named exponential regression equation.

Observe the following points:

(a) The regression curves (13) and (14) intersect at the point

$$u = \gamma + \frac{\lambda cr}{2}$$

$$v = be^{\frac{c^2}{2}} + a = m_v .$$

(b) When $r = 0$, the curves become two straight lines:

$$\bar{u} = \gamma = m_u$$

$$\bar{v} = be^{\frac{c^2}{2}} + a = m_v$$

which show that $\bar{u}$ is independent of $v$ and $\bar{v}$ is independent of $u$. We can also see this from the expression $F(u,v)$, which becomes

$$F(u,v) = \frac{1}{\sqrt{2\pi}\lambda} e^{-\frac{1}{2}\left(\frac{u-\gamma}{\lambda}\right)^2} \frac{1}{\sqrt{2\pi}\,c(v-a)} e^{-\frac{1}{2}\left(\frac{1}{c}\log\frac{v-a}{b}\right)^2}$$

when $r = 0$. This is the condition for independence of $u$ and $v$ in a probability sense.

(c) When $r = 1$, these two regression curves coincide. This signifies that there exists a complete functional relationship between $u$ and $v$, namely:

$$\frac{u-\gamma}{\lambda} = \frac{1}{c}\log\frac{v-a}{b} .$$

(d) As we have learned from the studies on the normal correlation surface, $r$ is the coefficient of correlation measuring the linear relationship between $x=\frac{u-\gamma}{\lambda}$ and $y=\frac{1}{c}\log(v-a)$. Thus, it is also a measure of relationships (13) and (14) existing between $u$ and $v$. If we note that $r$ may be written as

$$r=\frac{\mu_{11}}{\sigma_u\sigma_v}\,\frac{(e^{c^2}-1)^{\frac{1}{2}}}{c}$$

$$=\frac{\mu_{11}}{\sigma_u\sigma_v}\left(1+\frac{c^2}{2!}+\frac{c^4}{3!}+\frac{c^6}{4!}+\cdots\right)^{\frac{1}{2}} \tag{15}$$

we see that $r$ is always greater than $\mu_{11}/\sigma_u\sigma_v$, which would be the coefficient of correlation measuring the linear relationship between $u$ and $v$, if we treated the correlation surface of $u$ and $v$ as being normal.

The smaller the value of $c$ and $\alpha_{03}$, the smaller the difference between $r$ and $\mu_{11}/\sigma_u\sigma_v$. In fact, we can show, as we did for one variable case, that, as $c$ goes to zero the semi-logarithmic correlation surface approaches the normal correlation surface.

Incidentally, we may remark that the expression (15) is convenient for computing $r$.

## 4. *REGRESSION OF THE MOMENTS*

Using the well-known formulae for the moments of the normal curve of error about the mean, we can find at once the $s$-$th$ moment of $\theta_1(u,v)$ about its mean:

$$M_{s\cdot u}=\int_{-\infty}^{\infty}(u-\bar{u})^s\theta_1(u,v)\,du \tag{16}$$

$$=\frac{s!}{2^{s/2}\left(\frac{s}{2}\right)!}\,\lambda^s(1-r^2)^{s/2}, \qquad \text{if } s \text{ is even}$$

$$=0, \qquad \text{if } s \text{ is odd.}$$

This is the regression equation of the $s$-$th$ moment of $u$ about

the mean on $v$. It follows that the $s^{th}$ standard moment of $u$ for a given value of $v$ is:

$$\alpha_{s:u} = \frac{M_{s.u}}{M_{2.u}^{s/2}}$$

$$= \frac{s!}{2^{s/2}\left(\frac{s}{2}\right)!}, \qquad \text{if } s \text{ is even} \tag{17}$$

$$= 0, \qquad \text{if } s \text{ is odd.}$$

Again, by the formulae given in Part I for the moments of the logarithmic distribution, we calculate the $s^{th}$ moment of $\theta_2(u,v)$ about the point "$a$":

$$M'_{s \cdot v} = \int_a^{\infty} (v-a)^s \, \theta_2(u,v)\, dv$$
$$= b^s e^{scr\frac{u-\gamma}{\lambda} + \frac{s^2c^2(1-r^2)}{2}} . \tag{18}$$

And the regression equation of the $s^{th}$ moment of $v$ about the mean on $u$ is:

$$M_{s \cdot v} = \int_a^{\infty} (v-\bar{v})^s \, \theta_2(u,v)\, dv$$

$$= b^s e^{scr\frac{u-\gamma}{\lambda} + \frac{sc^2(1-r^2)}{2}} \sum_{k=0}^{s} (-1)^{s-k} \binom{s}{k} e^{\frac{k(k-1)c^2(1-r^2)}{2}} . \tag{19}$$

The $s^{th}$ standard moment of $v$ for a particular value of $u$ is, therefore,

$$\alpha_{s:v} = \frac{M_{s:v}}{M_{2 \cdot v}^{s/2}}$$

$$= \frac{\sum_{k=0}^{s} (-1)^{s-k} \binom{s}{k} e^{\frac{k(k-1)c^2(1-r^2)}{2}}}{\left(e^{c^2(1-r^2)} - 1\right)^{s/2}} . \tag{20}$$

Having obtained the expressions for the regressions of the moments of one variable on the other, we shall now proceed to

discuss the scedasticity, clisy and synagic* of the semi-logarithmic correlation surface.

## 5. SCEDASTICITY

From formula (16), we have the regression of the second moment of $u$ about the mean on $v$.

$$M_{2\cdot u} = \lambda^2(1-r^2) \tag{21}$$

which is the same as in the case of the normal correlation surface, except that $r$ now does not measure the linear relationship between $u$ and $v$. Since (21) is free of $v$, the semi-logarithmic correlation surface is homoscedastic, so far as the variable $u$ is concerned.

From the standpoint of estimation, we may also interpret expression (21) to mean that when we estimate the mean value of $u$ for a particular value of $v$, the error of estimation will be reduced if we use formula (13) instead of the mean of the marginal distribution of $u$. The standard deviation of the marginal distribution of $u$ is $\lambda$, while that of (13) is only $\sqrt{M_{2\cdot u}} = \lambda\sqrt{1-r^2}$ as shown by (21).

The second moment of $v$ for a particular value of $u$ is given by (19):

$$M_{2\cdot v} = b^2 e^{2cr\frac{u-h}{\lambda} + c^2(1-r^2)}\left[e^{c^2(1-r^2)} - 1\right] \tag{22}$$

which is not independent of $u$. So, the semi-logarithmic correlation surface is not homoscedastic for $v$. Actually $\sqrt{M_{2\cdot v}}$, the standard deviation of the distribution of $v$ for a given $u$, increases with $u$.

However, the relative dispersion or relative error for the

---

* The term "synagic" was used by S. D. Wicksell to mean the regression of the kurtosis. ("The Correlation Function of Type A, and the Regression of its Characteristics", Kungl. Svenska Vetenskapsakademiens Handlingar, Band 58, Nr. 3; Meddelanden fran Lunds Observatorium, Ser. II, Nr. 17, 1917)

distributions of $v$ for different values of $u$ is a constant, namely:

$$D_v = \frac{M}{\bar{v}-a} = \left[e^{c^2(1-r^2)}-1\right]^{\frac{1}{2}} . \tag{23}$$

Thus, by using **formula (14) to estimate the mean value** of $v$ for a given value of $u$ instead of employing the mean of the marginal distribution of $v$, we reduce the relative error of estimation, for the relative error of the marginal distribution is $(e^{c^2}-1)^{\frac{1}{2}}$. The reduction of relative error is much pronounced when $r$ is large. In fact, the greater $r$ is, the greater the reduction of relative error and the better the estimation. Hence, $r$ measures the degree of relationships (13) and (14) between $u$ and $v$.

## 6. CLISY AND SYNAGIC

Now, we shall study the clisy and synagic of the semi-logarithmic correlation surface or the regression of the skewness and kurtosis of one variable on the other.

The skewness and kurtosis of any distribution represented by $\theta_1(u,v)$ as measured by $\alpha_{3,4}$ and $\eta_u = \alpha_{4,u}-3$ are, of course, equal to zero, since it is a normal distribution. But the skewness and kurtosis of any distribution of $v$ for particular values of $u$, according to formula (20), are given by:

$$\alpha_{3,v} = (e^{c^2(1-r^2)}-1)^{\frac{1}{2}}(e^{c^2(1-r^2)}+2) \tag{24}$$

$$\eta_v = (e^{c^2(1-r^2)}-1)(e^{3c^2(1-r^2)}+3e^{2c^2(1-r^2)}+6e^{c^2(1-r^2)}+6) \tag{25}$$

which are two constants. Since the skewness and kurtosis of the marginal distribution of $v$ are given by $(e^{c^2}-1)^{\frac{1}{2}}(e^{c^2}+2)$ and $(e^{c^2}-1)(e^{3c^2}+3e^{2c^2}+6e^{c^2}+6)$, respectively, we may say that

the distribution of $v$ for each array of $u$ has smaller skewness and kurtosis, and is, therefore, closer to the normal distribution than the marginal distribution of $v$. And it is more so, when $r$ is near unity.

## 7. REGRESSION OF OTHER CHARACTERISTICS

In this section, we shall give the regression of other characteristics, such as the median, the geometric mean, the mode, the points of inflection and the finite limit.

The regression equation of the median and the mode of $u$ on $v$ are, of course, the same as that of the mean of $u$ on $v$, because $\theta_1(u,v)$ is normal. The points of inflection of $\theta_1(u,v)$ are points one standard deviation, i.e., $\sqrt{M_{2:u}}$, to the left and the right of the mean, as this is again a well-known property of the normal distribution.

The regression equation of the median and the geometric mean of $v$ on $u$ is given by

$$m_{d:v} = m_{g:v} = be^{cr\frac{u-\gamma}{\lambda}} + a$$

$$\text{or} \quad \frac{1}{c}\log\left(\frac{m_{d.v}-a}{b}\right) = r\frac{u-\gamma}{\lambda} \tag{26}$$

which differs from the regression equation of the mean or the median of $u$ on $v$, only in that the constant factor $r$ is on the left member of equation (13) but is on the right member of (26).

The mode of $v$ for special values of $u$ is

$$m_{o,v} = be^{cr\left(\frac{u-\gamma}{\lambda}\right)-c^2(1-r^2)} + a. \tag{27}$$

The regression equations of the points of inflection of $v$ on $u$ are given by

$$\breve{z}_{1,v} = be^{cr\frac{u-\gamma}{\lambda}-\frac{3}{2}c^2(1-r^2)\left[1+\sqrt{1+\frac{4}{9c^2(1-r^2)}}\right]} + a$$

$$\breve{z}_{2:v} = be^{cr\frac{u-\gamma}{\lambda}-\frac{3}{2}c^2(1-r^2)\left[1-\sqrt{1+\frac{4}{9c^2(1-r^2)}}\right]} + a$$

which are not free of $u$.

Finally, we may add that the finite limit of any distribution of $v$ for a particular array of $u$ is the same as that of the marginal distribution of $v$.

## 8. *AN ILLUSTRATION*

For illustrating the application of the semi-logarithmic correlation surface, we take the correlation table of heights and weights of 11,382 school boys between 5 and 14 years of age in Glasgow from L. Isserlis's paper, "On the Partial Correlation Ratio".* We shall treat the height as the variable $u$ and the weight as the variable $v$. Thus, the marginal distribution of heights is supposed to be normal, while that of weights is supposed to be logarithmic.

Letting the class marks, 49 inches and 56 pounds, be the origins of $u$ and $v$, respectively, and the class intervals be the respective units, we calculate the moments of this correlation surface:**

| | |
|---|---|
| $m_u = -\ 511861$ | class intervals |
| $\sigma_u = 1.7631$ | class intervals |
| $\alpha_{30} = .0177$ | |
| $\alpha_{40} = 2.5093$ | |
| $m_v = -.205412$ | class intervals |
| $\sigma_v = 2\ 5781$ | class intervals |
| $\alpha_{03} = .5915$ | |
| $\alpha_{04} = 3.1221$ | |
| $\mu_{11} = 4.205875$ | |

from which we deduce the following parameters by formulae (10):

| | |
|---|---|
| $\gamma = -\ 511861$ | class intervals |
| $\lambda = 1.7631$ | class intervals |
| $\omega = 1.0379$ | |
| $c = .1929$ | |
| $a = -13.45$ | class intervals |
| $b = 13.00$ | class intervals |
| $r = .9340$ | |

TABLE VI

Correlation Table of Heights and Weights of 11,382 School Boys between 5 and 14 Years of Age in Glasgow

(Original Measurements of Heights Made to Nearest Inch; Original Measurements of Weights Made to Nearest Pound)

Height (Inches)

| Weight (Pounds) Class Limit | 30-32 | 33-35 | 36-38 | 39-41 | 42-44 | 45-47 | 48-50 | 51-53 | 54-56 | 57-59 | 60-62 | 63-65 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24- 28 | 4 | 9 | 2 | | | 1 | | | | | | | 16 |
| 29- 33 | 3 | 42 | 62 | 25 | 3 | 1 | | | | | | | 136 |
| 34- 38 | | 16 | 220 | 414 | 72 | 6 | | | | | | | 728 |
| 39- 43 | 1 | 3 | 51 | 617 | 697 | 95 | 11 | 1 | | | | | 1476 |
| 44- 48 | | 1 | 7 | 122 | 875 | 603 | 38 | 8 | 1 | | | | 1655 |
| 49- 53 | | | 4 | 12 | 249 | 988 | 411 | 33 | 5 | 4 | | | 1706 |
| 54- 58 | | 1 | 3 | 1 | 17 | 436 | 905 | 171 | 11 | 4 | 3 | | 1552 |
| 59- 63 | | | 1 | | 1 | 39 | 630 | 568 | 51 | 6 | 1 | | 1297 |
| 64- 68 | | | | 1 | | 8 | 161 | 621 | 206 | 3 | 2 | 2 | 1004 |
| 69- 73 | | | | 1 | | | 35 | 374 | 340 | 24 | 2 | | 776 |
| 74- 78 | | | | | | | 3 | 106 | 335 | 76 | 5 | | 525 |
| 79- 83 | | | | | | | 2 | 22 | 120 | 93 | 4 | 1 | 242 |
| 84- 88 | | | | | | 1 | | 8 | 32 | 87 | 8 | 2 | 138 |
| 89- 93 | | | | | | | | 1 | 10 | 36 | 18 | 1 | 66 |
| 94- 98 | | | | | | | | | 3 | 23 | 9 | 2 | 37 |
| 99-103 | | | | | | | | | | 5 | 11 | 3 | 19 |
| 104-108 | | | | | | | | | 1 | | 5 | 1 | 7 |
| 109-113 | | | | | | | | | | | 1 | | 1 |
| 114-118 | | | | | | | | | | | | | 0 |
| 119-123 | | | | | | | | | | | | 1 | 1 |
| Total | 8 | 72 | 350 | 1193 | 1914 | 2178 | 2196 | 1913 | 1115 | 361 | 69 | 13 | 11,382 |

With these parameters, the correlation surface of heights and weights is determined. Now, we shall examine the regression curves of this correlation surface

Inserting the computed parameters in formulae (13) and (14), we obtain the regression equations of the mean height on weight and the mean weight on height. In Tables VII and VIII, we have the mean heights for specified weights and the mean weights for specified heights, We see, from these tables and from figures III and IV, the agreement the theoretical and observed results is very excellent In some extreme classes the deviations of the observed values from the theoretical values are more pronounced. But these classes comprise only a small fraction of the total number of cases.

Now, we go further to investigate the scedasticity of the correlation surface of heights and weights. According to the theory, for any particular weight the standard deviation of heights should be a constant and equal to $\sigma(1-r^2)^{\frac{1}{2}} = 1.8893$ inches. This is much less than the standard deviation of the marginal distribution of the heights, which is 5.2893 inches. That 1.8893 inches is quite close to the observed standard deviations is shown by Table IX and Figure V.

The theory asserts that the dispersion of weights is not the same for different heights. But for all arrays of heights the relative dispersion or relative error of weights is independent of heights.

TABLE VII

The Mean Heights for Specified Weights

| | Mean Height (Inches) | |
|---|---|---|
| Weight (Pounds) | Observed | Theoretical |
| 24- 28 | 34.4 | 33 2 |
| 29- 33 | 36.5 | 36.4 |
| 34- 38 | 39.3 | 39.3 |
| 39- 43 | 41.8 | 41.9 |
| 44- 48 | 44.0 | 44.2 |
| 49- 53 | 46.4 | 46.4 |
| 54- 58 | 48.5 | 48.3 |
| 59- 63 | 50.5 | 50.2 |
| 64- 68 | 52.1 | 51.9 |
| 69- 73 | 53.2 | 53.5 |
| 74- 78 | 54.9 | 55.0 |
| 79- 83 | 56.0 | 56.4 |
| 84- 88 | 57.1 | 57.8 |
| 89- 93 | 58.4 | 59.1 |
| 94- 98 | 58.8 | 60.3 |
| 99-103 | 60.7 | 61.5 |
| 104-108 | 60.6 | 62.6 |
| 109-113 | 61.0 | 63.6 |
| 114-118 | .... | 64.7 |
| 119-123 | 63.0 | 65.7 |

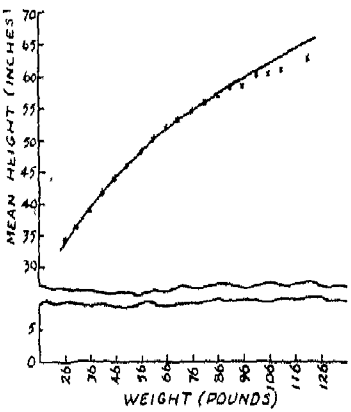


FIGURE III
REGRESSION CURVE OF MEAN HEIGHT ON WEIGHT

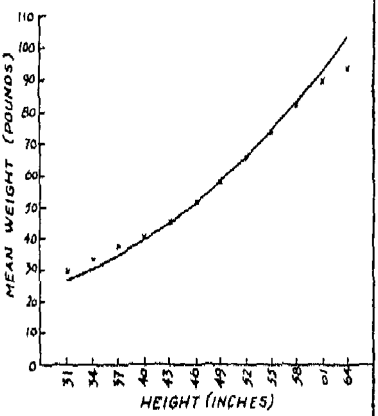


FIGURE IV
REGRESSION CURVE OF MEAN WEIGHT ON HEIGHT

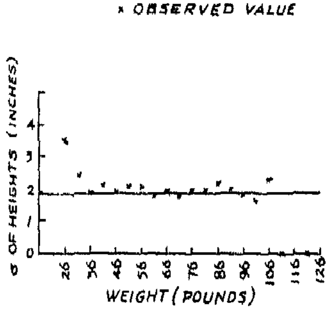


FIGURE V
CURVE OF SCEDASTICITY OF HEIGHT ON WEIGHT

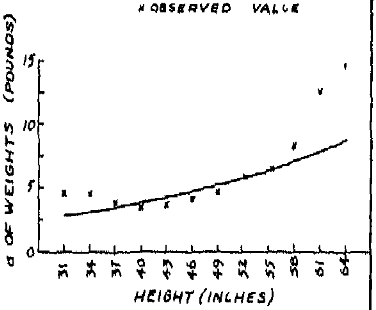


FIGURE VI
CURVE OF SCEDASTICITY OF WEIGHT ON HEIGHT

TABLE VIII

The Mean Weights for Specified Heights

| Height (Inches) | Mean Weight (Pounds) Observed | Theoretical |
|---|---|---|
| 30-32 | 29 8 | 26 0 |
| 33-35 | 32 5 | 30.0 |
| 36-38 | 36.4 | 34.4 |
| 39-41 | 39.7 | 39 3 |
| 42-44 | 44.6 | 44.7 |
| 45-47 | 50.4 | 50.8 |
| 48-50 | 57 3 | 57.5 |
| 51-53 | 65 1 | 64 8 |
| 54-56 | 72.6 | 73.1 |
| 57-59 | 81 7 | 82.1 |
| 60-62 | 89.1 | 92.2 |
| 63-65 | 92.2 | 103.3 |

According to formula (23), for any specified height, the relative error of weights is 7.6%, which is much smaller than the relative error of the marginal distribution of weights, which is

$$(e^{c^2}-1)^{\frac{1}{2}} = 19.5\%;$$

Both the theoretical and observed absolute errors or standard deviations of weights for specified heights have been calculated and are shown in Table X and Figure VI. The agreement between the theoretical and observed dispersions is not as good as for the regression of the mean weight on height. It should be noted here that theoretically the standard deviations of weights for heights over 76 inches are greater than the standard deviation of the marginal distribution of weights, which is 12.8905 pounds.

In interpreting the standard deviations of weights for particular heights, we must bear in mind that the distribution of weights for any given height is not normal, but logarithmic. Hence, a proper interpretation of the dispersion of weights for a given height can be made only with reference to the skewness, masured by the third standard moment of weights, which, according to the theory, is a constant for all different heights. The theoretical third standard moment of the distribution of weights for any given height, as we shall see later, is approximately .2.

TABLE IX

The Standard Deviations of Heights for Specified Weights

| | Standard Deviation of Heights (Inches) | |
|---|---|---|
| Weight (Pounds) | Observed | Theoretical |
| 24- 28 | 3 52 | 1 89 |
| 29- 33 | 2.40 | 1 89 |
| 34- 38 | 1.91 | 1.89 |
| 39- 43 | 2.12 | 1.89 |
| 44- 48 | 1.91 | 1.89 |
| 49- 53 | 2.07 | 1.89 |
| 54- 58 | 2.04 | 1.89 |
| 59- 63 | 1 81 | 1 89 |
| 64- 68 | 1 87 | 1 89 |
| 69- 73 | 1.79 | 1.89 |
| 74- 78 | 1 92 | 1.89 |
| 79- 83 | 1.95 | 1.89 |
| 84- 88 | 2 18 | 1 89 |
| 89- 93 | 2 01 | 1.89 |
| 94- 98 | 1 86 | 1 89 |
| 99-103 | 1 62 | 1.89 |
| 104-108 | 2 34 | 1.89 |
| 109-113 | 0 | 1.89 |
| 114-118 | .... | 1.89 |
| 119-123 | 0 | 1.89 |

TABLE X

The Standard Deviations of Weights for Specified Heights

| | Standard Deviation of Weights (Pounds) | |
|---|---|---|
| Height (Inches) | Observed | Theoretical |
| 30-32 | 4.6 | 2 8 |
| 33-35 | 4 5 | 3.1 |
| 36-38 | 4.0 | 3.5 |
| 39-41 | 3.5 | 3.8 |
| 42-44 | 3.6 | 4.3 |
| 45-47 | 4.2 | 4.7 |
| 48-50 | 4.8 | 5.2 |
| 51-53 | 5.9 | 5.8 |
| 54-56 | 6.3 | 6.4 |
| 57-59 | 8 4 | 7.1 |
| 60-62 | 12.5 | 7.9 |
| 63-65 | 14 8 | 8.7 |

Thus, from Table II in Part I, we find that the probability that any weight will be at most one standard deviation above or below the mean weight for a given height is .6839 instead of .6826, as in the case of the normal distribution. The difference between .6839 and .6826 is slight but should not be overlooked. Moreover, the difference would not be so small, if the skewness were larger.

Another thing we must observe is that since the standard deviation of weights for a given height increases with height, the probability that for a given height the weight will differ from the mean weight for that height by, say, at most one pound is not the same for all different heights, although the probability that for a given height the weight will differ from the mean weight for that height by at most one standard deviation is the same for all different heights. The former probability is greater for smaller heights.

The agreement between the theoretical and observed clisy and synagic is, of course, not expected to be close. Theoretically, the distributions of weights for specified heights should all have $\alpha_{3\cdot v} = .23$ and $\eta_v = \alpha_{4:v} - 3 = .09$. Five observed values of $\alpha_{3:v}$ and $\eta_v$ are shown below:

| Height (Inches) | Observed Skewness of Weights $\alpha_{3\cdot v}$ | Observed Kurtosis of Weights $\eta_v$ |
|---|---|---|
| 36-38 | .22 | 8.72 |
| 42-44 | .19 | .18 |
| 48-50 | .29 | .79 |
| 54-56 | .12 | 1.54 |
| 60-62 | -.93 | .50 |

The rather large deviations of the observed $\eta_\nu$ in the first class from the theoretical $\eta_\nu$ and the observed $\alpha_{3\cdot\nu}$ in the last class from the theoretical $\alpha_{3\cdot\nu}$ may be accounted for by the fact that only 350 and 69 observations are included in the first and the last classes, respectively

The observed marginal distribution of heights is very symmetric but is markedly leptokurtic, since its $\alpha_4$ is about 2.5093. Hence, the fit given by a normal curve is not quite satisfactory, as we can see from Table XI.

The observed marginal distribution of weights is quite skew and platykurtic. As shown by Table XII, the agreement between the observed distribution and the theoretical logarithmic distribution is not very close

TABLE XI

Relative Frequency Distribution of Heights of 11,382 School Boys between 4 and 15 Years of Age in Glasgow

| Class Limits (Inches) | Observed Relative Frequency | Theoretical Relative Frequency (Normal Curve) |
|---|---|---|
| 27-29 | | .0003 |
| 30-32 | .0007 | .0020 |
| 33-35 | .0063 | 0095 |
| 36-38 | .0308 | .0332 |
| 39-41 | 1048 | 0846 |
| 42-44 | .1682 | 1577 |
| 45-47 | .1913 | .2154 |
| 48-50 | .1929 | 2143 |
| 51-53 | .1681 | .1561 |
| 54-56 | .0980 | .0831 |
| 57-59 | .0317 | .0324 |
| 60-62 | .0061 | .0092 |
| 63-65 | 0011 | .0019 |
| 66-69 | | .0003 |
| Total | 1.0000 | 1.0000 |

TABLE XII

Relative Frequency Distribution of Weights of 11,382 School Boys between 4 and 15 Years of Age in Glasgow

| Class Limits (Pounds) | Observed Relative Frequency | Theoretical Relative Frequency (Logarithmic Curve) |
|---|---|---|
| 19- 23 | | .0006 |
| 24- 28 | 0014 | .0048 |
| 29- 33 | .0119 | .0211 |
| 34- 38 | 0640 | 0564 |
| 39- 43 | .1297 | .1041 |
| 44- 48 | .1454 | .1441 |
| 49- 53 | .1499 | 1609 |
| 54- 58 | .1363 | .1504 |
| 59- 63 | .1139 | .1234 |
| 64- 68 | .0882 | .0897 |
| 69- 73 | 0682 | .0602 |
| 74- 78 | 0461 | 0373 |
| 79- 83 | .0213 | .0212 |
| 84- 88 | 0121 | 0127 |
| 89- 93 | .0058 | .0065 |
| 94- 98 | 0033 | 0033 |
| 99-103 | .0017 | .0018 |
| 104-108 | 0006 | .0008 |
| 109-113 | .0001 | .0004 |
| 114-118 | | .0002 |
| 119-123 | .0001 | .0001 |
| Total | 1.0000 | 1.0000 |

In closing, we may say that the semi-logarithmic correlation surface is not at all uncommon in practice, and the method developed here for treating it should prove rather useful. In fact, our investigation opens up a new way for determining exponential and logarithmic regression curves.

## BIBLIOGRAPHY

This Bibliography contains the names of the books and articles which deal with the logarithmic frequency distribution, as well as those which are related to the general problem of the transformation of frequency distributions. But no attempt has been made to include all the literature concerning the transformation of frequency distributions.

ANDERSON, WALTER—Researches into the Theory of Regression, pp. 126-148 and pp. 153-154 (Meddelande från Lunds Astronomiska Observatorium, Ser. II, Nr. 64, 1932)

BAKER, G. A.—Transformations of Bimodal Distributions (Annals of Mathematical Statistics, Vol I, No. 4, 1930, pp. 334-44)

BERNSTEIN, S—Sur les Courbes de Distribution des Probabilités (Mathematische Zeitschrift, Vol 24, 1925, pp 199-211)

BRUNS, H.—Wahrscheinlichkeitsrechnung und Kollektivmasslehre, pp. 126-139 (B. G. Teubner, Leipzig and Berlin, 1906)

CHARLIER, C V L.—Das Strahlungsgesetz (Meddelande från Lunds Astronomiska Observatorium, No 55, 1913 or Arkiv för Matematik, Astronomi och Fysik, Band 9, No. 11)
Contributions to the Mathematical Theory of Statistics, 7. Frequency Curves of Compound Functions (Meddelande från Lunds Astronomiska Observatorium, No. 61, 1914)

DAVIES, GEORGE R—The Logarithmic Curve of Distribution (Journal of the American Statistical Association, Vol XX, 1925, pp. 467-480)
The Analysis of Frequency Distributions (Journal of the American Statistical Association, Vol XXIV, 1929, pp 349-366)

DODD, EDWARD L.—The Frequency Law of a Function of Variables with Given Frequency Laws (Annals of Mathematics, Second Series, Vol. 27, 1926, pp 12-20)

EDGEWORTH, E. F.—On Representation of Statistics by Mathematics (J. R. S. S., Vol. LXI, 1898, pp. 670-700)
A Method of Representing Statistics by Analytic Geometry (Proceedings of the Fifth International Congress of Mathematicians, Vol. II, 1912, pp 427-40)
On the Use of Analytic Geometry to Represent Certain Kinds of Statistics (J. R. S. S., Vol. LXXVII, 1914, pp. 300-312, pp 415-432, pp 653-671, pp. 838-852)
On Mathematical Representation of Statistical Data (J. R. S. S., Vol LXXIX, 1916, pp. 455-500; Vol LXXX, 1917, pp. 65-83, pp. 266-288, pp. 411-437)

FISHER, ARNE*—The Mathematical Theory of Probabilities, pp 235-260 (The MacMillan Co., New York, 1928)

GALTON, FRANCIS—The Geometric Mean in Vital and Social Statistics, (Proc. of the Royal Society, Vol. 29, 1879, pp. 365-367)

JENKINS, THOMAS N.—A Short Method and Tables for the Calculation of the Average and the Standard Deviation of Logarithmic Distribution (Annals of Mathematical Statistics, Vol. III, No. 1, 1932, pp. 45-55)

*Mistakes have been found in the results given by Arne Fisher and Jorgensen.

JØRGENSEN, N. R.*—Undersøgelser over Frequensflader og Korrelation, pp 45-51 (Arnold Busck København, 1916)

KAPTEYN, J. C.—Skew Frequency Curves in Biology and Statistics, (Noordhoff, Groningen, 1903)
Reply to Professor Pearson's Criticisms (Recueil des Travaux Botanique Néarlandais, Vol. 2, 1906, pp. 216-222)
Skew Frequency Curves in Biology and Statistics, 2nd paper, Ch. I (Hoitsema Brothers, Groningen, 1916)

MCALLISTER, DONALD—The Law of the Geometric Mean (Proc of the Royal Society, Vol 29, 1879, pp 367-376)

NYDELL, STURE—The Mean Errors of the Characteristics in the Logarithmic-Normal Distribution (Skandinavisk Aktuarietidskrift, 1919, pp 134-144)

PEARSON, KARL—"Das Fehlergesetz und Seine Verallgemeinerunger durch Fechner und Pearson", a Rejoinder (Biometrika, Vol IV, 1905, pp 170-212)
Skew Frequency Curves, A Rejoinder to Professor Kapteyn (Biometrika, Vol V, 1906 pp. 168-171)

Pretorius, S J.—Skew Bivariate Frequency Surfaces Examined in the Light of Numerical Illustrations (Biometrika, Vol. XXII, 1930, pp 109-223)

RIETZ, H L.- Frequency Distributions Obtained by Certain Transformations of Normally Distributed Variates (Annals of Mathematics, Second Series, 1923, pp 292-300)
On Certain Properties of Frequency Distributions of the Powers and Roots of Variates of a Given Distribution (Proc of the National Academy, Vol. 13, No. 12, 1927, pp. 817-820)
On Certain Properties of Frequency Distributions Obtained by a Linear Fractional Transformation of the Variates of a Given Distribution (Annals of Math Statistics, Vol II, No 1, 1931, pp. 38-47)

VAN UVEN, M. J.—Skew Frequency Curves in Biology and Statistics, 2nd Paper, Chs II and III (Hoitsema Brothers, Groningen, 1916)
Logarithmic Frequency Distribution (Proc. Koninklijke Akademie van Wetenschappen, Amsterdam, Vol XIX, No. 3, 1917, pp 533-546)
Skew Frequency Curves (Proc Kon Ak v. Wet, Amsterdam, Vol. XIX, No. 3, 1917, pp. 670-684)
On Treating Skew Correlation (Proc Kon. Ak. v. Wet., Amsterdam, Vol. XXVIII, Nos 8-9, 1925, pp. 797-811; No 10, pp. 919-935; Vol. XXIX, No. 4, 1926, pp 580-590; Vol. XXXII, No. 4, 1929, pp. 408-413)
Skew Correlation between Three and More Variables (Proc Kon. Ak. v. Wet, Amsterdam, Vol XXXII, No 6, 1929, pp. 793-807, No 7, pp 995-1007, No. 8, pp 1085-1103)

WICKSELL, S D.—On the Genetic Theory of Frequency (Arkiv för Matematik, Astronomi och Fysik, Band 12, No. 20, 1917)
On Logarithmic Correlation with an Application to the Distribution of Ages at First Marriage (Meddelande från Lunds Astronomiska Observatorium, No 84, 1917)
Das Heiratsalter in Schweden 1891-1910, Ein Korrelationsstatistische Untersuchung (Lunds Universitets Årsskrift, N. F. Avd. 2, Bd. 14, Nr. 18, 1918 or Kungl. Fysiografiska Sallskpets Handlingar, N. F Bd. 29, Nr. 18)

---

*Mistakes have been found in the results given by Arne Fisher and Jørgensen.

# A SIMPLE METHOD FOR CALCULATING MEAN SQUARE CONTINGENCY

By

ELMER B. ROYER
The Ohio State University

*If we wish to test for a possible relationship between two* variables which are not quantitatively measurable, but each of which has two or more categories, the usual procedure is to make a two-way table, giving the frequencies of all the possible combinations.

Assuming independence between the two variables, a second table is built, making the frequencies of each column proportional to the frequencies in the column of row totals. When this is done, each of the row frequencies is found to be proportional to the row of column totals.

The deviation of the actual frequency for a compartment as found in Table 1 from the expected frequency as found in Table 2, is squared and this square is divided by the expected frequency. These quotients are summed over the entire table, giving us Chi-square.

The calculation of Chi-square can be made much simpler by simplifying the formula.

The probability of the occurrence of two independent events is the product of their separate probabilities. Thus the probability of the joint occurrence of Category 3 of the first classification and Category d of the second classification is the probability of the occurrence Category 3 (which is taken to be the fraction of the total number of cases which fall in Category 3), times the probability of the occurrence of Category d. The expected frequency of the compartment is this product of separate probabilities, multiplied by the total number of cases. If we let $f_a$ be the actual compartment frequency, $f_e$ the expected compartment frequency,

$f_r$ the total frequency for the row, $f_c$ the total frequency for the column, and $N$ the number of cases, we may write,

$$\begin{aligned} f_e &= \frac{f_r}{N} \cdot \frac{f_c}{N} \cdot N \\ &= \frac{f_r\, f_c}{N} \end{aligned} \tag{1}$$

Also,

$$\begin{aligned} \chi^2 &= \Sigma \frac{(f_a - f_e)^2}{f_e} \\ &= \Sigma \frac{f_a^2 - 2 f_a f_e + f_e^2}{f_e} \\ &= \Sigma \frac{f_a^2}{f_e} - 2\Sigma f_a + \Sigma f_e . \end{aligned} \tag{2}$$

Since the table must sum to $N$, whether we have filled it with actual frequencies, or with expected frequencies, (2) reduces to,

$$\chi^2 = \Sigma \frac{f_a^2}{f_e} - N$$

Substituting for $f_e$ from (1),

$$\begin{aligned} \chi^2 &= \Sigma \frac{f_a^2}{\frac{f_r f_c}{N}} - N \\ &= N \Sigma \frac{f_a^2}{f_r f_c} - N \\ &= N \left[ \Sigma \frac{f_a^2}{f_r f_c} - 1 \right] \end{aligned} \tag{3}$$

In order to illustrate our method, we shall compute Chi-square for Table 18 on page 86 of Fisher's *Statistical Methods for Research Workers*. Our computations are presented in the following table

| | Black Self | Black Piebald | Brown Self | Brown Piebald | Total |
|---|---|---|---|---|---|
| Coupling F₁ Males | 88<br>7744<br>25.3902 | 82<br>6724<br>22.0459 | 75<br>5625<br>18.4426 | 60<br>3600<br>11.8033 | 305 |
| F₁ Females | 38<br>1444<br>11.7398 | 34<br>1156<br>9.3984 | 30<br>900<br>7.3171 | 21<br>441<br>3.5854 | 123 |
| Repulsion F₁ Males | 115<br>13225<br>31.6388 | 93<br>8649<br>20.6914 | 80<br>6400<br>15.3110 | 130<br>16900<br>40.4306 | 418 |
| F₁ Females | 96<br>9216<br>25.7430 | 88<br>7744<br>21.6313 | 95<br>9025<br>25.2095 | 79<br>6241<br>17.4330 | 358 |
| Frequency Total | 337 | 297 | 280 | 290 | 1204 |
| Product Total | 94.5118 | 73.7670 | 66.2802 | 73.2523 | |
| Quotient | .280450 | .248374 | .236715 | .252594 | 1.018133 |

The actual frequency is the first entry in each compartment. The square is read from a table and written directly beneath the frequency. The reciprocal of 305 is put into the keyboard of a calculator and multiplied in turn by each of the squares in the first row. The products make the third entries in the compartments.

These products are summed by columns and the sums divided by the frequency totals of the corresponding columns. These quotients are summed horizontally. This sum is $\Sigma \dfrac{f_a^2}{f_r f_c}$, and can be substituted in (3). For our example,

$$X^2 = 1204(1.018133 - 1.000000)$$
$$= 21\ 832$$

This answer agrees exactly with the answer obtained by Fisher in his Table 19, page 87. The advantages of this method are two-fold: (1) There is considerable saving of labor; (2) with the simplification of calculations, we have greatly reduced the danger of errors caused by dropping of decimal places.

*THE AMERICAN STATISTICAL ASSOCIATION*

Vol. IV No. 2

# THE ANNALS

of

# MATHEMATICAL STATISTICS

*SIX DOLLARS PER ANNUM*

MAY 1933

PUBLISHED QUARTERLY BY

AMERICAN STATISTICAL ASSOCIATION

*Publication Office*—Edwards Brothers, Inc., Ann Arbor, Michigan

*Business Office*—530 Commerce Bldg., New York Univ., New York, N. Y.

*Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.*

STATEMENT OF OWNERSHIP

UNDER ACT OF CONGRESS OF AUGUST 24, 1912

*Publisher*—American Statistical Association, New York City, New York.

*Editor*—H. C. Carver, University of Michigan.

*Managing Editor*—Rattan S. Sekhon, Ann Arbor, Michigan.

*Business Manager*—J. W. Edwards, Ann Arbor, Michigan.

*Owner*—American Statistical Association, 530 Commerce Bldg., New York City.

# A METHOD OF DETERMINING THE CONSTANTS IN THE BIMODAL FOURTH DEGREE EXPONENTIAL FUNCTION

*By*

A. L. O'Toole

In a paper in this Journal[1] the present writer has discussed some of the mathematical properties of a class of definite integrals which arise in the study of the frequency function

$$y = e^{-a^2(x^4+p_1x^3+p_2x^2+p_3x+p_4)}, \quad a \neq 0. \tag{1}$$

This function defines the system of frequency curves for which the method of moments is the best method of fitting[2]—i.e. best in the sense of maximum likelihood—and this fact gives importance to its study. The curves are typically bimodal, the nature and location of the modes being given by the roots of the equation

$$4x^3+3p_1x^2+2p_2x+p_3 = 0. \tag{2}$$

The first problem which arose was that of finding an expression for the value of the definite integral

$$I_0 = \int_{-\infty}^{\infty} e^{-a^2(x^4+p_1x^3+p_2x^2+p_3x+p_4)}\,dx. \tag{3}$$

If $x$ is replaced by $x - \frac{p_1}{4}$ this integral becomes

$$I_0 = \int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx+r)}\,dx, \tag{4}$$

---

[1] On the system of curves for which the method of moments is the best method of fitting. Vol. IV, No. 1, Feb. 1933, p. 1.

[2] R. A. Fisher, On the mathematical foundations of theoretical statistics, Philosophical Transactions of the Royal Society of London, vol. 222, series A (1921), p. 355.

or

$$(5) \quad I_0 = k \int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)} dx \quad \text{where} \quad k = e^{-a^2 r},$$

or, replacing $x\sqrt{a}$ by $x$ where $a$ is the positive square root of $a^2$

$$(6) \qquad I_0 = \frac{k}{\sqrt{a}} \int_{-\infty}^{\infty} e^{-(x^4+apx^2+a^{\frac{3}{2}}qx)} dx,$$

or

$$(7) \qquad I_0 = K \int_{-\infty}^{\infty} e^{-(x^4+Px^2+Qx)} dx,$$

where $\qquad K = \frac{k}{\sqrt{a}}, \quad P = ap, \quad Q = a^{\frac{3}{2}}q.$

No real loss of generality is incurred in studying (5), (6) or (7) rather than (3). For the purposes of the previous paper it was found convenient to discuss certain special cases of (7) first, then (7) itself and later (5). Having in mind the practical purposes of this note, however, attention will be focused first on the form (5) and afterwards on (3). The transformations from the expressions obtained in the previous paper are very simple. For (5) the special cases studied and a few of the more important results obtained may be stated here as follows:

Type I:

$p = q = 0.$

$$I_0 = k \int_{-\infty}^{\infty} e^{-a^2x^4} dx = \frac{k}{\sqrt{2a}} \Gamma\left(\tfrac{1}{4}\right),$$

$$I_{2n} = k \int_{-\infty}^{\infty} x^{2n} e^{-a^2x^4} dx = \frac{k}{2a^{(2n+1)/2}} \Gamma\left(\frac{2n+1}{4}\right), \quad n = 0, 1, 2, 3, \ldots,$$

$$I_{2n+1} = k \int_{-\infty}^{\infty} x^{2n+1} e^{-a^2x^4} dx = 0, \quad n = 0, 1, 2, 3, \cdots\cdots,$$

$$u_1 = \frac{I_1}{I_0} = 0,$$

$$u_2 = \frac{I_2}{I_0} = \frac{\Gamma(\frac{3}{4})}{a\Gamma(\frac{1}{4})},$$

$$u_3 = \frac{I_3}{I_0} = 0,$$

$$u_4 = \frac{I_4}{I_0} = \frac{\Gamma(\frac{5}{4})}{a^2\Gamma(\frac{1}{4})} = \frac{1}{4a^2},$$

hence

$$a^2 = \frac{1}{4u_4},$$

$$u_{2n} = \frac{I_{2n}}{I_0} = \frac{\Gamma(\frac{2n+1}{4})}{a^n\Gamma(\frac{1}{4})}, \quad n = 0, 1, 2, 3, \ldots\ldots,$$

$$u_{2n+1} = \frac{I_{2n+1}}{I_0} = 0, \quad n = 0, 1, 2, 3, \ldots\ldots .$$

Obviously, of course, $k$ depends upon the total frequency and hence if the total frequency is

$$k = \frac{N}{\int_{-\infty}^{\infty} e^{-a^2x^4}\,dx} = \frac{2N\sqrt{a}}{\Gamma(\frac{1}{4})}.$$

This curve has a single mode located at $x = 0$ and is symmetrical with respect to the ordinate at $x = 0$.

Type II:

$$q=0,\ p=-2b,\ b>0,$$

$$I_0=k\int_{-\infty}^{\infty} e^{-a^2(x^4-2bx^2)}dx$$

$$=\frac{k}{\sqrt{a}}\int_{-\infty}^{\infty} e^{-(x^4-2abx^2)}dx$$

$$=\frac{k}{\sqrt{a}}\Big[\tfrac{1}{2}\Gamma(\tfrac{1}{4})(1+\frac{a^2b^2}{2!}+\frac{5\cdot 1a^4b^4}{4!}+\frac{9\cdot 5\cdot 1a^6b^6}{6!}+\frac{13\cdot 9\cdot 5\cdot 1a^8b^8}{8!}+\ldots)$$
$$+ab\Gamma(\tfrac{3}{4})(1+\frac{3a^2b^2}{3!}+\frac{7\cdot 3\cdot a^4b^4}{5!}+\frac{11\cdot 7\cdot 3a^6b^6}{7!}+\ldots)\Big]$$

$$=\frac{k}{\sqrt{a}}\Big[e^{\frac{a^2b^2}{2}}\tfrac{1}{2}\Gamma(\tfrac{1}{4})(1+\frac{a^4b^4}{3\cdot 4}+\frac{a^8b^8}{7\cdot 3\cdot 4^2\cdot 2!}+\ldots)$$
$$+ab\Gamma(\tfrac{3}{4})(1+\frac{a^4b^4}{5\cdot 4}+\frac{a^8b^8}{9\cdot 5\cdot 4^2\, 2!}+\ldots)\Big]$$

It was shown that this integral could be expressed in terms of the Bessel functions $J_{\frac{1}{4}}$ and $J_{-\frac{1}{4}}$ as follows:

$$I_0=\frac{k}{\sqrt{a}}\left[(\frac{a^2b^2}{2})^{\frac{1}{4}}e^{\frac{a^2b^2}{2}}\left\{AJ_{\frac{1}{4}}(\frac{-ia^2b^2}{2})+BJ_{-\frac{1}{4}}(\frac{-ia^2b^2}{2})\right\}\right]$$

where

$$A=\frac{2^{\frac{3}{4}}\Gamma(\frac{1}{4})\Gamma(\frac{3}{4})}{\sqrt[4]{-i}},$$

$$B=\frac{\frac{1}{2}\Gamma(\frac{1}{4})\Gamma(\frac{3}{4})}{\sqrt[4]{2i}},\quad i=\sqrt{-1}.$$

If the total frequency is $N$

$$k = \frac{N}{\int_{-\infty}^{\infty} e^{-a^2(x^4-2bx^2)}\,dx}$$

This curve is symmetrical with respect to the ordinate at $x=0$ and has two real modes located at $x=\pm\sqrt{b}$.

Type III: $p=0$.

$$I_0 = k\int_{-\infty}^{\infty} e^{-a^2(x^4+qx)}\,dx$$

$$=\frac{k}{\sqrt{a}}\int_{-\infty}^{\infty} e^{-(x^4+a^{\frac{3}{2}}qx)}\,dx$$

$$=\frac{k}{2\sqrt{a}}\sum_{n=0}^{\infty}\frac{(a^{\frac{3}{2}}q)^{2n}}{(2n)!}\,\Gamma\left(\frac{2n+1}{4}\right)$$

$$=\frac{k}{2\sqrt{a}}\left[\Gamma\left(\tfrac{1}{4}\right)\left\{1+\frac{(a^{\frac{3}{2}}q)^4}{4\cdot 4!}+\frac{5(a^{\frac{3}{2}}q)^8}{4^2\cdot 8!}+\frac{9\cdot 5(a^{\frac{3}{2}}q)^{12}}{4^3\cdot 12!}+\cdots\right\}\right.$$
$$\left.+\Gamma\left(\tfrac{3}{4}\right)\left\{\frac{(a^{\frac{3}{2}}q)^2}{2!}+\frac{3(a^{\frac{3}{2}}q)^6}{4.6!}+\frac{7\cdot 3(a^{\frac{3}{2}}q)^{10}}{4^2\cdot 10!}+\cdots\right\}\right],$$

$$k = \frac{N}{\int_{-\infty}^{\infty} e^{-a^2(x^4+qx)}\,dx}$$

This curve is not symmetrical and has only one real mode, that mode being located at $x$ equal to the real cube root of negative $q$.

Type IV: The general case.

$$I_0 = k\int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)}dx$$

$$= \frac{k}{a}\int_{-\infty}^{\infty} e^{-(x^4+apx^2+a^{\frac{3}{2}}qx)}dx$$

$$= K\int_{-\infty}^{\infty} e^{-(x^4+Px^2+Qx)}dx.$$

It was shown that the value of this integral could be expressed as an infinite series each term of which involved two Bessel functions. But, as pointed out near the close of the previous paper, although this infinite series may be considered a theoretical solution of the problem, it does not lead to a simple method of determining the constants $a^2, p, q, k$ which appear in the frequency function. It is the purpose of this note to give a practical method of determining these constants.

Beginning with (5)

$$I_0 = k\int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)}dx,$$

the $n$-$th$ moment $u_n'$ is defined by

(8) $$u_n' = \frac{k\int_{-\infty}^{\infty} x^n e^{-a^2(x^4+px^2+qx)}dx}{k\int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)}dx}$$

$$= \frac{k}{I_0}\int_{-\infty}^{\infty} x^n e^{-a^2(x^4+px^2+qx)}dx.$$

Integrate $I_0$ by parts, letting $u = e^{-a^2(x^4+px^2)}$ and $dv = e^{-a^2qx}dx$. Then

$$(9) \qquad I_0 = -\frac{k}{q}\int_{-\infty}^{\infty}(4x^3+2px)e^{-a^2(x^4+px^2+qx)}dx.$$

Divide by $I_0$ and multiply by $q$ and the result is

$$(10) \qquad q = -(4u_3' + 2pu_1').$$

Start again with $I_0$ in the form (5) and integrate by parts letting

$u = e^{-a^2(x^4+px^2+qx)}$ and $dv = dx$. Then

$$(11) \qquad I_0 = ka^2\int_{-\infty}^{\infty}(4x^4+2px^2+qx)e^{-a^2(x^4+px^2+qx)}dx.$$

Divide by $I_0$ and then

$$1 = a^2(4u_4' + 2pu_2' + qu_1')$$

or

$$(12) \qquad a^2 = \frac{1}{4u_4' + 2pu_2' + qu_1'}.$$

Now integrate (11) by parts with $u = e^{-a^2(x^4+px^2+qx)}$ and

$dv = (4x^4 + 2px^2 + qx)\,dx$. This leads to

$$(13) \qquad \begin{aligned} I_0 = \frac{ka^4}{30}\int_{-\infty}^{\infty}&(96x^8 + 128px^6 + 84qx^5 + 40p^2x^4 \\ &+ 50pqx^3 + 15q^2x^2)e^{-a^2(x^4+px^2+qx)}dx. \end{aligned}$$

Divide by $I_0$ and obtain

$$1 = \frac{a^4}{30}(96u_8' + 128pu_6' + 84qu_5' + 40p^2u_4' + 50pqu_3' + 15q^2u_2')$$

or

$$(14)\quad a^4 = \frac{30}{96u_8' + 128pu_6' + 84qu_5' + 40p^2u_4' + 50pqu_3' + 15q^2u_2'}.$$

Squaring in (12) the result is

$$(15)\quad a^4 = \frac{1}{(4u_4' + 2pu_2' + qu_1')^2}$$

$$= \frac{1}{16u_4'^2 + 4p^2u_2'^2 + q^2u_1'^2 + 16pu_2'u_4' + 8qu_1'u_4' + 4pqu_1'u_2'}.$$

Eliminating $a^4$ between (14) and (15) the equation

$$(16)\quad p^2(40u_4' - 120u_2'^2) + q^2(15u_2' - 30u_1'^2) + pq(50u_3' - 120u_1'u_2')$$
$$+ p(128u_6' - 480u_2'u_4') + q(84u_5' - 240u_1'u_4') + (96u_8' - 48u_4'^2) = 0.$$

Using relation (10) and

$$(17)\qquad q^2 = 16u_3'^2 + 16pu_1'u_3' + 4p^2u_1'^2$$

eliminate $q$ from (16) obtaining

$$(18)\qquad 5Ap^2 + 2Bp + 2C = 0$$

and hence

$$(19)\qquad p = \frac{-B \pm \sqrt{B^2 - 10AC}}{5A}$$

where

$$(20)\begin{cases} A = 2u_4' - 6u_2' + 15u_1'^2u_2' - 6u_1'^4 - 5u_1'u_3', \\ B = 90u_1'u_2'u_3' - 60u_1'^3u_3' - 25u_3'^2 + 16u_6' - 60u_2'u_4' - 21u_1'u_5' + 60u_1'^2u_4', \\ C = 30u_2'u_3'^2 - 60u_1'^2u_3'^2 - 42u_3'u_5' + 120u_1'u_3'u_4' + 12u_8' - 60u_4'^2. \end{cases}$$

In order to decide upon one of the two values of $p$ furnished by (18) notice that, equating the first derivative of the frequency function to zero, the location of the two modes and the minimum point between them is determined by the roots of the equation

$$4x^3+2px+q=0. \tag{21}$$

The condition for three real distinct roots in this equation is

(22) $-8p^3>27q^2$, which requires $p<0$,

where $q$ is found from (17). If $-8p^3=27q^2$ then one of the modes coincides with the minimum point. If $p=q=0$ then both modes coincide with the minimum point.

Extracting the square root in (17) gives two values of $q$ differing only in sign. Now it is easy to show either by geometrical considerations or by examining the algebraic manipulations leading to (18) that $p$ is independent of the sign of $q$. Changing $q$ to $-q$ in (5) has the same effect as changing $x$ to $-x$ or, that is, reversing the order of the distribution and curve. Also, changing $x$ to $-x$ leaves the even moments unaltered but changes the sign of every odd moment. Hence if the value of the function at the modal position on the left is greater than the value of the function at the modal position on the right then $q$ is greater than zero. And if the value of the function at the modal position on the left is less than the value of the function at the modal position on the right then $q$ is less than zero. If $q=0$ the curve is symmetrical with respect to the ordinate at $x=0$. Hence $p$ and $q$ are determined by (19), (17) and (22), the sign of $q$ being fixed by examination of the data of the problem or, if necessary, by trial. The value of $a^2$ is then found by taking the positive square root in (15). Of course (14) would give the same value for $a^2$.

Now that $a^2$, $p$ and $q$ are determined, there remains only $k$ to be found. If the total frequency is $N$ then

$$k\int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)}dx = N$$

and hence

$$k = \frac{N}{\int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx)}dx} \tag{23}$$

where the numerical value of the integral in the denominator can be found by mechanical quadrature to any desired degree of approximation. For purposes of the quadrature involved here it will be found that the simple rectangle quadrature formula will give as good results as could be desired.[8] Having found $k$ then the constant $r$ is also known since

$$\left[\begin{array}{l} k = e^{-a^2 r}, \\ \\ r = \dfrac{\log_e k}{-a^2} = \dfrac{\log_{10} k}{-a^2 \log_{10} e} \end{array}\right. \tag{24}$$

The points of inflexion are located by equating the second derivative of the function to zero. The equation is

$$a^2(4x^3+2px+q)^2-2(6x^2+p)=0. \tag{25}$$

If now $x$ be replaced by $x+m$ then

$$I_0 = \int_{-\infty}^{\infty} e^{-a^2(x^4+px^2+qx+r)}dx \tag{5}$$

becomes

$$I_0 = \int_{-\infty}^{\infty} e^{-a^2(x^4+p_1x^3+p_2x^2+p_3x+p_4)}dx \tag{3}$$

[8]On the degree of Approximation of Certain Quadrature Formulas. Annals of Mathematical Statistics, vol. IV, No. 2, May 1933, p. 143 by A. L. O'Toole.

where

$$(26)\qquad \begin{cases} p_1 = 4m \\ p_2 = 6m^2+p, \\ p_3 = 4m^3+2mp+q, \\ p_4 = m^4+m^2p+mq+r. \end{cases}$$

The data[4] in the first two columns of the table given here will provide the basis for an illustration of the method described above for determining the constants. The numbers in the first column are the classes into which the plants were divided. In the second column are found the frequencies corresponding to the various classes. In constructing the third column the origin for $x$ was arbitrarily placed to correspond to the class 25. Taking

$$u'_n = \frac{\Sigma x^n f(x)}{\Sigma f(x)}$$

the first six moments and the eighth moment are found to be

$$u'_1 = \frac{88}{452} = 0.1946903,$$

$$u'_2 = \frac{14086}{452} = 31.16372,$$

$$u'_3 = \frac{12248}{452} = 27.09735,$$

$$u'_4 = \frac{1000264}{452} = 2212.973,$$

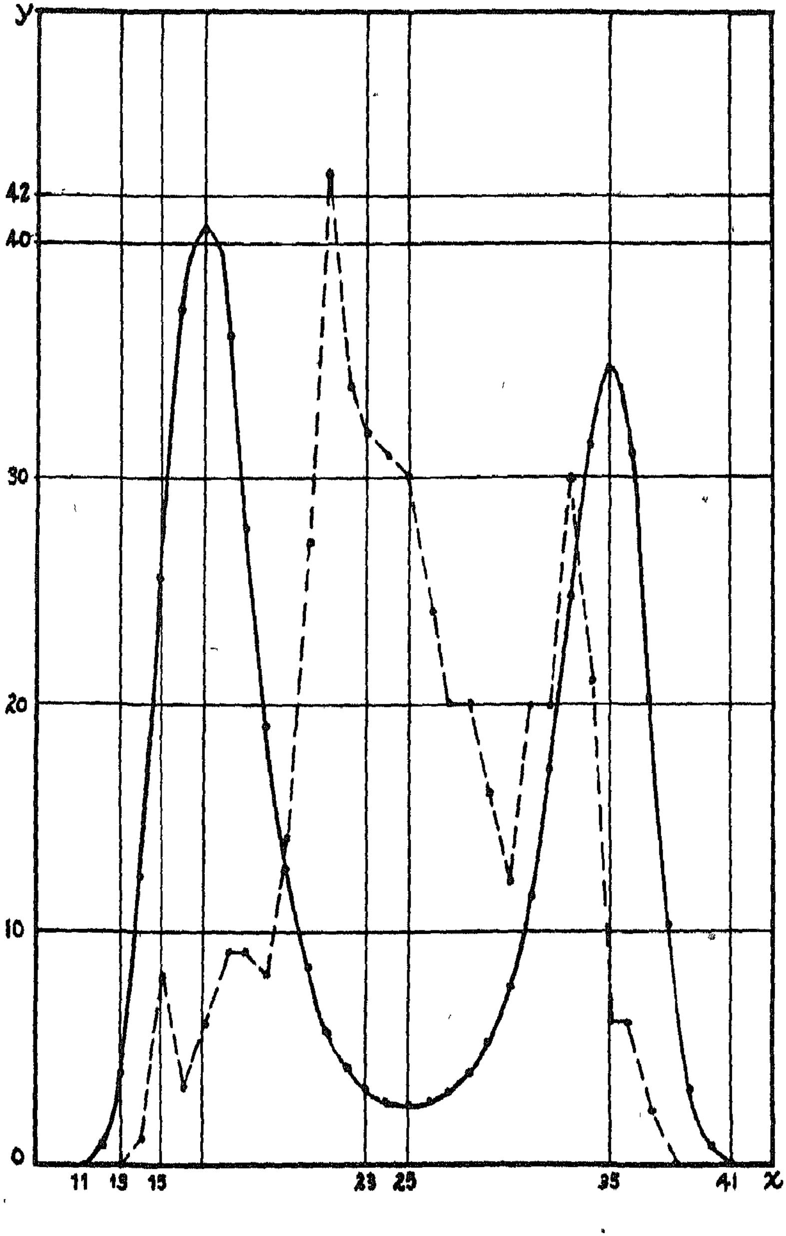
$$u'_5 = \frac{185480}{452} = 410.3540,$$

$$u'_6 = \frac{94571296}{452} = 209228.5,$$

$$u'_8 = \frac{10428472504}{452} = 23071842.$$

[4]This data, except for slight modifications, was extracted from that of W. L. Tower on the Seriation of Counts of Rays of Chrysanthemum Leucanthemum, Biometrika No. 1, 1901-2, p. 313.

| CLASS | FREQUENCY $f(x)$ | $x$ | $y'$ | $2.5y'=y$ | COMPUTED FREQUENCY |
|---|---|---|---|---|---|
| 9 | | -16 | .031102 | .077755 | 0 |
| 10 | | -15 | .300472 | .751180 | 1 |
| 11 | | -14 | 1.583594 | 3.958985 | 4 |
| 12 | 1 | -13 | 4.991068 | 12.477670 | 12 |
| 13 | 8 | -12 | 10.246676 | 25.616690 | 26 |
| 14 | 3 | -11 | 14.831798 | 37.079495 | 37 |
| 15 | 6 | -10 | 16.280112 | 40.700280 | 41 |
| 16 | 9 | - 9 | 14.482540 | 36.206350 | 36 |
| 17 | 9 | - 8 | 11.089036 | 27.722590 | 28 |
| 18 | 8 | - 7 | 7.712195 | 19.280487 | 19 |
| 19 | 14 | - 6 | 5.108903 | 12.772257 | 13 |
| 20 | 27 | - 5 | 3.359072 | 8.397680 | 8 |
| 21 | 43 | - 4 | 2.269766 | 5.674415 | 6 |
| 22 | 34 | - 3 | 1.621764 | 4.054410 | 4 |
| 23 | 32 | - 2 | 1.252760 | 3.131900 | 3 |
| 24 | 31 | - 1 | 1.062910 | 2.657275 | 3 |
| 25 | 30 | 0 | 1.000000 | 2.500000 | 3 |
| 26 | 24 | 1 | 1.046530 | 2.616325 | 3 |
| 27 | 20 | 2 | 1.214445 | 3.036112 | 3 |
| 28 | 20 | 3 | 1.547935 | 3.869837 | 4 |
| 29 | 16 | 4 | 2.133051 | 5.332627 | 5 |
| 30 | 12 | 5 | 3.108096 | 7.770240 | 8 |
| 31 | 20 | 6 | 4.654336 | 11.635840 | 12 |
| 32 | 20 | 7 | 6.917724 | 17.294310 | 17 |
| 33 | 30 | 8 | 9.793409 | 24.483522 | 24 |
| 34 | 21 | 9 | 12.593310 | 31.483275 | 31 |
| 35 | 6 | 10 | 13.938151 | 34.845377 | 35 |
| 36 | 6 | 11 | 12.502565 | 31.256412 | 31 |
| 37 | 2 | 12 | 8.504388 | 21.260970 | 21 |
| 38 | | 13 | 4.078577 | 10.196442 | 10 |
| 39 | | 14 | 1.274131 | 3.185327 | 3 |
| 40 | | 15 | .238029 | .595072 | 1 |
| 41 | | 16 | .024259 | .060647 | 0 |
| | 452 | | 180.792704 | | 452 |


y
42
40
30
20
10
0
11
13
15
23
25
35
41
x

Formulas (20) give

$$A = -1409.786$$

$$B = -790428.9,$$

$$C = -15354106.$$

Hence from (19)

$$p = -202.7862$$

or

$$p = -21.48292.$$

But $p = -21.48292$ and the value of $q$ to which it leads do not satisfy the relation (22) hence use $p = -202.7862$. Calculate $q$ from (17) and use the positive square root since an examination of the data shows that the value of the function at the left modal value is greater than the value of the function at the right modal value. Hence

$$q = 29.4284.$$

Formula (15) now gives as the positive square root

$$a^2 = 0.0002640.$$

Using these values for $a^2, p, q$ the values of the function

$$y' = e^{-a^2(x^4 + px^2 + qx)}$$

are calculated for integral values of $x$ from $x = -16$ to $x = 16$ and tabulated in column four. The constant $k$ is then found by dividing the total frequency 452 by the sum of column four. Hence

$$k = \frac{452}{180.792704} = 2.500100.$$

By (24)

$$r = -3472.578.$$

The function can now be written

$$y = 2.5e^{-.0002640(x^4 - 202.7862x^2 + 29.4284x)} \text{ taking } k = 2.5,$$

or $y = e^{-.0002640(x^4 - 202.7862x^2 + 29.4284x - 3472.578)}$.

The values of the ordinates for this function are given in column five and to the nearest integer in column six.

Equation (21) becomes

$$4x^3-2(202.7862)x+29.4284=0$$

which has the roots (approximate) $x=-10.1$, $x=0.07$, $x=10.03$. It should be noted that the sum of the three roots must equal zero. Hence the modes are located at $x=-10.1$ and at $x=10.03$ with the minimum point at $x=0.03$. These roots can be determined to any desired number of decimal places by Horner's method.

If now $x$ is replaced by $x=-25$ so that the new values of $x$ are respectively equal to the numbers in the class column, the function becomes

$$y=e^{-.0002640(x^4-100x^3+3547.214x^2-52331.26x+259565.3)}$$

The modes are now located at $x=14.9$ and $x=35.03$ with the minimum point at $x=25.07$. In the figure are shown the original distribution and the curve represented by this equation.

# ON THE TCHEBYCHEF INEQUALITY OF BERNSTEIN

*By*

CECIL C. CRAIG[1]

From Tchebychef's inequality we know that if $x_1, x_2, \cdots, x_n$ are a set of independent statistical variables with

$$m_{x_1} = m_{x_2} = \cdots\cdots = m_{x_n} = 0,$$

and

$$\sigma^2 = \sigma^2_{x_1} + \sigma^2_{x_2} + \cdots\cdots + \sigma^2_{x_n},$$

then the probability $P$ that

$$-t\sigma \leq x_1 + x_2 + \cdots + x_n \leq t\sigma$$

satisfies the inequality,

$$P \geq 1 - \frac{1}{t^2} .$$

This gives a lower limit for $P$ which is often unsatisfactory. Improvement of this result requires further hypotheses. As is well-known, Pearson, Camp, Guldberg, Meidel, Narumi,[2] and Smith[3] have attacked this problem with considerable success. Another interesting and important attempt in this direction due to S. Bernstein seems to have generally escaped attention in the English-speaking world, at least, since it has been published only in Russian.[4] Because of the latter fact, it seems necessary to give

[1]This paper was written in substantially its present form during the author's tenure of a National Research Fellowship at Stanford University.

[2]For references to all these papers except Smith's and a brief discussion see Rietz, H. L., Mathematical Statistics, (Open Court Publishing Company, Chicago, 1927), pp. 140-144.

[3]Smith, C. D., On Generalized Tchebychef Inequalities in Mathematical Statistics, American Journal of Mathematics, Vol. 52, (1930), pp. 109-126.

[4]Bernstein, S., Theory of Probability, (Moscow, 1927), pp. 159-165. The present account of this work of Bernstein is taken from a lecture of Professor J. V. Uspensky.

a brief account of this work of Bernstein's preliminary to the remarks based on it the writer wishes to make.

Bernstein imposed the condition in addition that

$$(1)\quad E(|x_i|^k) \leq \frac{\sigma_{x_i}^2}{2} k!\, h^{k-2};\ k \geq 2,\quad i = 1, 2, \cdots, n,$$

($E(x)$ is read "the mathematical expectation of $x$.") in which $h$ is arbitrary. (This condition is satisfied, e.g., if the $x_i$'s are bounded.) and used the following lemma due to Tchebychef. Let the statistical variable $u$ be always $> 0$. If $E(u) = A$, then the probability $Q$ that $u \geq At^2$ satisfies the inequality, $Q \leq \frac{1}{t^2}$

Then taking,

$$u = e^{\varepsilon(x_1 + x_2 + \cdots\cdots + x_n)},$$

$$= e^{\varepsilon x_1} e^{\varepsilon x_2} \cdots\cdots e^{\varepsilon x_n},$$

in which $\varepsilon$ is arbitrary,

$$E(u) = E(e^{\varepsilon x_1}) E(e^{\varepsilon x_2}) \cdots\cdots E(e^{\varepsilon x_n}).$$

Now

$$e^{\varepsilon x_i} = 1 + \varepsilon x_i + \frac{\varepsilon^2 x_i^2}{2!} + \frac{\varepsilon^3 x_i^3}{3!} + \cdots,$$

and under the condition (1),

$$E(e^{\varepsilon x_i}) \leq 1 + \frac{\varepsilon^2 \sigma_{x_i}^2}{2} + \frac{\varepsilon^3 \sigma_{x_i}^2 h}{2} + \frac{\varepsilon^4 \sigma_{x_i}^2 h^2}{2} + \cdots$$

If it is assumed that

$$|\varepsilon| h \leq c < 1$$

then

$$E(e^{\varepsilon x_i}) \leq 1 + \frac{\varepsilon^2 \sigma_{x_i}^2}{2(1-c)} < e^{\frac{\varepsilon^2 \sigma_{x_i}^2}{2(1-c)}},$$

and thus

$$(2)\qquad E(u) < e^{\frac{\varepsilon^2 \sigma^2}{2(1-c)}}$$

If in the inequality, $u \geq At^2$, a greater quantity is substituted

for $A$, then certainly $Q \leq \frac{1}{t^2}$. Therefore the probability $Q$ of

$$u \geq e^{\frac{\mathcal{E}^2\sigma^2}{2(1-c)}} e^{\tau^2}$$

satisfies the inequality

$$Q \leq e^{-\tau^2}$$

Now

$$u = e^{\mathcal{E}(x_1 + x_2 + \cdots x_n)} \geq e^{\tau^2 + \frac{\mathcal{E}^2\sigma^2}{2(1-c)}}$$

implies for $\mathcal{E} > 0$,

$$x_1 + x_2 + \cdots + x_n \geq \frac{\tau^2}{\mathcal{E}} + \frac{\mathcal{E}\sigma^2}{2(1-c)}.$$

The value of $\mathcal{E}$ is next chosen so as to make $Q$ a minimum, i.e., so as to make $\frac{\tau^2}{\mathcal{E}} + \frac{\mathcal{E}\sigma^2}{2(1-c)}$ a minimum. Thus

$$\mathcal{E}^2 = \frac{2(1-c)\tau^2}{\sigma^2}.$$

Then the probability $Q$ that

$$x_1 + x_2 + \cdots + x_n \geq \tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}}$$

satisfies the inequality,

$$Q \leq e^{-\tau^2}$$

if $\mathcal{E}^2 = \frac{2(1-c)\tau^2}{\sigma^2}$ ; $\mathcal{E} \leq \frac{c}{h}$ with $c < 1$.

To get the corresponding result for the lower limit of the sum $x_1 + x_2 + \cdots + x_n$, it is only necessary to choose $\mathcal{E} < 0$ and as before, the probability, $Q'$, that

$$x_1 + x_2 + \cdots + x_n \leq -\tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}}$$

satisfies the inequality,

$$Q' \leq e^{-\tau^2}$$

if also $\varepsilon^2 = \frac{2(1-c)}{\sigma^2}\tau^2$ and $|\varepsilon| \leq \frac{c}{h}$ with $c < 1$.

Combining these two results, if $P$ is the probability of

$$-\tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}} \leq x_1 + x_2 + \cdots + x_n \leq \tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}}$$

then since

$$P + Q + Q' = 1,$$

$$P \geq 1 - 2e^{-\tau^2}.$$

But setting

$$\tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}} = \omega,$$

and also,

$$\varepsilon^2 \leq \frac{c^2}{h^2},$$

the condition

$$\frac{2(1-c)}{\sigma^2}\tau^2 \leq \frac{c^2}{h^2}$$

(Bernstein set $\varepsilon^2 = \frac{c^2}{h^2}$, using merely the equality sign in this condition. The value of $c$ as here given is necessary in the author's developments below.) must be satisfied, or what is the same thing,

$$\frac{2(1-c)^2\omega^2}{2\sigma^4} \leq \frac{c^2}{h^2},$$

from which

$$c \geq \frac{h\omega}{\sigma^2 + h\omega}.$$

This last quantity on the right is positive and $< 1$ as required so that the constants can actually be chosen as specified.

This gives

$$\tau = \omega\left[2(\sigma^2 + h\omega)\right]^{-\frac{1}{2}},$$

and finally the probability, $P$, that

$$-\omega \leq x_1 + x_2 + \cdots x_n \leq \omega$$

is such that

$$P \geq 1 - 2e^{-\frac{\omega^2}{2\sigma^2 + 2h\omega}},$$

or setting $\omega = t\sigma$

$$P \geq 1 - 2e^{-\frac{t^2}{2 + \frac{2ht}{\sigma}}}. \tag{3}$$

It is to be observed that generally the quantity $\frac{2ht}{\sigma}$ rapidly decreases with increasing $n$.

This is the inequality reached by Bernstein under the condition (1).

If all the $x_i$'s are bounded, if, say, always

$$|x_i| \leq b, \quad i = 1, 2, \cdots\cdots, n,$$

one may take $h = \frac{b}{3}$.

It is the purpose of the author's remarks to discuss less severe conditions than (1) under which the inequality (3) can be obtained. These more general conditions are obtained, however, at the expense of assuming quite generally satisfied regularity conditions with regard to the "tails" of the frequency distribution of $x$, which needs not necessarily to be regarded as the sum of $n$ component variables, $x_1, x_2, \cdots, x_n$.

If we now take

$$u = e^{\varepsilon x} \tag{4}$$

we have

$$E(u) = \int_{-\infty}^{\infty} dF(x) e^{\varepsilon x} \quad (F(x) \text{ is the probability function of } x).$$

$$= \int_{-\infty}^{\infty} dF(x) \left(1 + \varepsilon x + \frac{\varepsilon^2 x^2}{2!} + \frac{\varepsilon^3 x^3}{3!} + \dots \right).$$

The condition (1) insures that the series under the sign of integration may be integrated over the interval $(-\infty, \infty)$. But the series can also be integrated over the same interval if it converges uniformly in any fixed finite interval, which it does, and if the series $\sum_{n=0}^{\infty} g_n(y)$, where

$$g_n(y) = \int_{-y}^{y} dF(x) \frac{\varepsilon^n x^n}{n!},$$

converges uniformly in the interval $(-\infty, \infty)$.

Formally, at least,

$$(5) \qquad E(u) = 1 + \mu_2 \frac{\varepsilon^2}{2!} + \mu_3 \frac{\varepsilon^3}{3!} + \dots ,$$

in which $\mu_k$ is the $k$-$th$ moment about the mean of $x$. If

$$(6) \qquad |\mu_k| \leq \frac{k!}{2} \mu_2 h^{k-2}, \; k \geq 2,$$

for some $h > 0$, then for $h|\varepsilon| \leq c < 1$ the right hand side of (5) is convergent and is $\leq 1 + \frac{\varepsilon^2 \sigma^2}{2(1-c)}$ as before. Now let us suppose that the condition (6) is satisfied not only for moments taken over the whole interval $(-\infty, \infty)$ but also for moments taken over any interval which includes the interval $(-b, b)$ in which $b$ is an arbitrarily large though finite number. This is the *first regularity condition*, mentioned above, which we shall impose on the tails of the frequency function of $x$.

Then it is obvious, from the remark above, that

$$\sum_{n=0}^{\infty} g_{n(y)}$$

is uniformly convergent in the interval for $|y| \leq b$ for $\frac{b}{3}|\mathcal{E}| \leq c < 1$ And for $|y| > b$ it is also obvious that for $h|\mathcal{E}| \leq c < 1$,

$$\sum_{n=0}^{\infty} g_n(y)$$

is uniformly convergent if our first regularity condition is satisfied. And since $|\mathcal{E}|$ may be taken arbitrarily small, the inequality (3) follows as before.

It is evident that if our first regularity condition holds, that the condition (6) is more general than the condition (1). And it is easily seen that this first regularity condition holds for a very wide class of frequency functions. For, in order for it to hold, it is sufficient that the frequency curve (continuous or not) outside some finite interval $(-b, b)$ about the mean as center, be never increasing as $|x|$ increases and that if $f(x)$ be the ordinate of the frequency curve at the abscissa $x$, always $f(x) \geq f(-x)$ or else always $f(x) \leq f(-x)$ for $x > b$.

But if the first regularity condition be satisfied, then for all intervals which include $(-b, b)$ the corresponding moments have upper limits in absolute value. And if this be so for all such intervals, the semi-invariants (of Thiele) will also have upper limits for their absolute values. If $\lambda_k$ is the $k$-$th$ semi-invariant, we will take for our *second regularity condition* on the tails of the frequency distribution of $x$, that

$$(7) \qquad |\lambda_k| \leq \frac{k!}{2} \lambda_2 h^{k-2} \quad k \geq 2 \quad (\lambda_2 = \mu_2)$$

for some $h > o$ if $\lambda_k$ is taken for any interval which includes the arbitrarily large, though finite, interval $(-b, b)$.

If this second regularity condition holds, it is again easy to show that (5) is an equality if $h|\mathcal{E}| \leq c < 1$. The right member

of (5) is still uniformly convergent in the interval $(-b, b)$ for $\frac{b}{3}\ |\mathcal{E}| \leq c < 1$. For all intervals which include $(-b, b)$ we use the formal identity which defines the semi-invariants of Thiele:

$$(8)\quad e^{\lambda_2 \frac{\mathcal{E}^2}{2!} + \frac{1}{3!}\lambda_3 \frac{\mathcal{E}^3}{3!} + \cdots}$$

$$= 1 + \mu_2 \frac{\mathcal{E}^2}{2!} + \mu_3 \frac{\mathcal{E}^3}{3!} + \cdots = e^{\phi(\mathcal{E})}$$

Under the condition (7), $\phi(\mathcal{E})$ is uniformly convergent over the intervals in question for $h|\mathcal{E}| \leq c < 1$ and for these values of $\mathcal{E}$, (8) becomes an equality since its second member is only the first arranged in powers of $\mathcal{E}$. Moreover, on account of (7) the right member must be uniformly convergent for all intervals which include $(-b, b)$.

At least one important class of frequency distributions satisfies our second regularity condition. The distributions of characteristics in samples of $N$ have finite ranges as long as $N$ is finite and they commonly have semi-invariants which are rapidly decreasing with increasing $N$. If such distributions approach normality their semi-invariants of order above the second approach zero, in particular they may become in absolute value less than or equal to the corresponding semi-invariants of a Pearson's Type III distribution which are given by

$$\frac{\lambda_k}{\lambda_2^{\frac{k}{2}}} = \frac{(k-1)!}{a^{k-2}} \qquad\qquad k \geq 2$$

in which $\quad a = \dfrac{2\lambda_2^{\frac{3}{2}}}{\lambda_3}\quad$, or

$$\lambda_k = (k-1)!\,\lambda_2 \left(\frac{\sigma}{a}\right)^{k-2}.$$

Taking $\ h = \left|\frac{\sigma}{a}\right|\ $ it is easy to see that such distributions satisf our second regularity condition. The smaller the skewness of th Type III distribution, the smaller $h$ may be taken. Thus in suc

cases we can give a lower limit for $P(|x|\le t\sigma)$, the probability that $|x|\le t\sigma$ , which is improved with decreasing skewness of the Type III distribution. By the use of the first regularity condition we could only take $h = \frac{\sigma}{2}$ as the distribution approaches normality.

As a second application, let us suppose that $x = x_1 + x_2$ in which $x_1$ and $x_2$ are independent, and in which the semi-invariants of the distribution of $x_1$ are $l_2\,(=\sigma_1^2),\ l_3,\ l_4 \cdots$, and the semi-invariants of the distribution of $x_2$ are $\gamma_2\,(=\sigma_2^2)\,\gamma_3,\ \gamma_4, \cdots$. Then the distribution of $x$ has for semi-invariants

$$\lambda_2 = l_2 + \gamma(=\sigma^2),\quad \lambda_3 = l_3 + \gamma_3,\quad \lambda_4 = l_4 + \gamma_4, \cdots\ .$$

Further let it be assumed that $\frac{\sigma_2}{\sigma_1} < 1$ , and that the distribution of $x_2$ satisfies our second regularity condition.

Then

$$P(|x|\le t\sigma) > P(|x_1|\le t\sigma_1)\, P(|x_2| \le t(\sigma-\sigma_1))$$

But

$$P(|x_2|\le t(\sigma-\sigma_1)) = P\left(|x_2| \le \frac{t(\sigma-\sigma_1)}{\sigma_2}\,\sigma_2\right)$$

$$> 1-2e^{-\frac{\frac{t^2(\sigma-\sigma_1)^2}{\sigma_2^2}}{2+2\frac{h}{\sigma_2}\,\frac{t(\sigma-\sigma_1)}{\sigma_2}}}\,.$$

Now

$$\frac{\sigma-\sigma_1}{\sigma_2} = \frac{(\sigma_1^2+\sigma_2^2)^{\frac{1}{2}}-\sigma_1}{\sigma_2} < 1 \qquad \left\{ \begin{array}{l} (1+x^2)^{\frac{1}{2}} < 1+x \\ \text{if } 0 < x < 1 \end{array}\right.$$

so that we get

$$P(|x_2| \le t(\sigma-\sigma_1)) > 1-2e^{-\frac{t^2}{2+2h\frac{t}{\sigma_2}}}\,.$$

This gives finally in such cases

$$P(|x|\le t\sigma) > P(|x_1|\le t\sigma)-2e^{-\frac{t^2}{2+2h\frac{t}{\sigma_2}}}$$

# ON CORRELATION SURFACES OF SUMS WITH A CERTAIN NUMBER OF RANDOM ELEMENTS IN COMMON*

*By*

CARL H. FISCHER

*Introduction.* The study of correlation due to a common factor has been a more or less familiar one in the literature of mathematical statistics. Kapteyn,[1] in an exposition of the Pearsonian coefficient of correlation, considered the correlation between two sums of normally distributed variables, the sums having $k$ random elements in common. In 1920, Rietz[2] devised urn schemata which yield sums with common items involved in such a way that the correlation and regression properties can be dealt by a priori methods. In a later paper, Rietz[3] considered two variables, each the sum of two random drawings of elements from a continuous rectangular distribution, with one of the elements in common. Here, the emphasis was placed principally upon the description of the correlation surface. Some other aspects and extensions of this problem were brought out by Karl Pearson[4] in an editorial discussion of Rietz's paper.

In the literature, the theory of correlation has been discussed principally in connection with its applications. One of the objects of some of the above-mentioned papers is the establishment of a closer connection between correlation theory and abstract probalibity theory. Such a connection would give a more precise

*Presented to the American Mathematical Society, Dec. 28, 1931.

[1] J. C. Kapteyn, "Definition of the Correlation-Coefficient," Monthly Notices of the Royal Astronomical Society, Vol. 72 (1912), pp. 518-525.

[2] H. L. Rietz, "Urn Schemata as a Basis for the Development of Correlation Theory," Annals of Mathematics, Vol. 21 (1920), pp. 306-322.

[3] H. L. Rietz, "A Simple Non-Normal Correlation Surface," Biometrika, Vol. 24 (1932), pp. 288-291.

[4] Karl Pearson, "Professor Rietz's Problem," (Editorial), Biometrika, Vol. 24 (1932), pp. 290-291.

meaning to correlation and would tend to make the study of correlation theory more attractive to mathematicians. With this aim in view, the present paper is concerned with correlation among sums having common elements, extending and generalizing the preceding papers in several ways.

We shall assume our drawings made from a continuous universe characterized by a rather arbitrary law of distribution. We shall define $n$ sums, each of an arbitrary number of elements, formed in such a manner that any two consecutive sums have elements in common, and inquire into the correlation between any two of these sums. The equations of the correlation surfaces will be expressed in terms of iterated integrals, the regression of each variable on the other will be shown to be linear, and the equations of the regression lines will be obtained. The coefficient of correlation may then be computed from the slopes of these lines.

Throughout this paper we shall understand a probability function, $f(t)$, to be, for all values of $t$ on a range $R$, a single-valued, real-valued, non-negative, continuous function of $t$. It is then Riemann integrable on $R$, and we shall require that $\int_R f(t)\,dt = 1$. We define the probability that a value of $t$, drawn at random from the range $R$, lie in the interval $(a, b)$, $a$ and $b$ in $R$ and $b > a$, to be $\int_a^b f(t)\,dt$. We may then say that $f(t)dt$ is, to within infinitesimals of higher order, the probability that a value of $t$ drawn at random lies in the interval $(t, t+\Delta t)$. Bachelier[5] has classified probabilities into those of the first, second, and third kinds, and Craig[6] has extended this to probability functions, according as $R$ is the range $(-\infty, \infty)$, $(O, \infty)$, and $(O, a)$, respectively. We shall find it convenient to adopt this classification.

---

[5] L. Bachelier, "Calcul des Probabilities." (1912), p. 155.

[6] Allen T. Craig, "On the Distribution of Certain Statistics," American Journal of Mathematics, Vol. 54 (1932), pp. 353-366.

I. Sums of elements drawn from a universe characterized by a probability function of the first kind.

1. The correlation between two sums having random elements in common. Let $f(t)$, a probability function of the first kind, characterize the distribution of the variable $t$. Let the principal variable $x_1$ be defined as the sum of $n_1$ independent values of $t$ drawn at random. Further, let the principal variable $x_2$ be defined as the sum of $k_{12}$ random values of the $n_1$ values of $t$ composing $x_1$ and of $n_2-k_{12}$ independent random values of $t$ taken directly from the universe characterized by $f(t)$.

Theorem I. *Given the sums $x_1$ and $x_2$ as defined above, with $k_{12}$ random elements in common.*

a) *The marginal distributions of $x_1$ and $x_2$ are given, respectively, by*

$$(1.11)\quad G_1(x_1)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(t_{11})f(t_{12})\cdots f(t_{1,n_1-1})f(x_1-t_{11}-\cdots-t_{1,n_1-1})dt_{1,n_1-1}\cdots dt_{11},$$

*and*

$$(1.12)\quad \begin{aligned}G_2(x_2)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}&f(t_{11})\cdots f(t_{1k_{12}})f(t_{2,k_{12}+1})\cdots f(t_{2,n_2-1})\\ &\times f(x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-1})dt_{2,n_2-1}\cdots dt_{2,k_{12}+1}dt_{1k_{12}}\cdots dt_{11}.\end{aligned}$$

b) *The correlation surface, $w=F(x_1,x_2)$, or the simultaneous law of distribution of $x_1$ and $x_2$, is given by*

$$(1.2)\quad \begin{aligned}F(x_1,x_2)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}&f(t_{11})\cdots f(t_{1,n_1-1})f(x_1-t_{11}-\cdots-t_{1,n_1-1})f(t_{2,k_{12}+1})\cdots f(t_{2,n_2-1})\\ &\times f(x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-1})dt_{2,n_2-1}\cdots dt_{2,k_{12}+1}dt_{1k_{12}}\cdots dt_{11}.\end{aligned}$$

c) *The regression curves of $x_2$ on $x_1$ and of $x_1$ on $x_2$ are linear, and are given, respectively, by the following equations:*

$$(1.31)\quad \bar{x}_2=\frac{k_{12}x_1}{n_1}+(n_2-k_{12})M,$$

*and*

$$\bar{x}_1 = \frac{k_{12}x_2}{n_2} + (n_1 - k_{12})M, \tag{1.32}$$

*where*

$$M = \int_{-\infty}^{\infty} t f(t)\,dt.$$

*Hence, the coëfficient of correlation between* $x_1$ *and* $x_2$ *is*

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{1/2}}.$$

*Proof.* The proof for the expressions for the marginal distributions of $x_1$ and $x_2$ are given by Craig[7] and need not be repeated here. The correlation surface $w = F(x_1, x_2)$ is derived by a simple extension of the same method to two independent variables.

The regression curve of $x_2$ on $x_1$ is the locus of the ordinate of the centroid $\bar{x}_2$ of a section of the surface for any given $x_1$. Thus

$$\bar{x}_2 = \frac{\left[\int_{-\infty}^{\infty} x_2 F(x_1, x_2)\,dx_2\right]}{\left[\int_{-\infty}^{\infty} F(x_1, x_2)\,dx_2\right]}. \tag{1.4}$$

It will be convenient in what follows to use an abbreviated notation by letting

$$\theta(x_1, t_{11}, \cdots t_{1,n_1-1}) = f(t_{11}) \cdots f(t_{1,n_1-1}) f(x_1 - t_{11} - \cdots - t_{1,n_1-1}), \tag{1.5}$$

which is merely the integrand of the marginal distribution of $x_1$. Where no ambiguity can result, $\theta_{x_1}$ will be used in place of $\theta(x_1, t_{11}, \cdots t_{1,n_1-1})$. Then $F(x_1, x_2)$ may be written

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta(x_1, t_{11}, \cdots t_{1,n_1-1}) \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1})$$

$$\times\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11}.$$

Now let $v = x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1}$. Changing the variable

[7] Allen T. Craig, loc. cit., pp. 355-356.

from $x_2$ to $v$, (1.4) becomes

$$(1.6) \quad \bar{x}_2 = \Big\{ \Big[ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{11} + \cdots + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{1k_{12}} + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{2,k_{12}+1} + \cdots$$

$$+ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{2,n_2-1} + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} v \Big] \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v) dt_{2,n_2-1} \cdots dt_{11} dv \Big\}$$

$$\Big/ \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v) dt_{2,n_2-1} \cdots dt_{11} dv \Big\}.$$

It will be noted that the terms in the numerator fall into two groups: those terms containing the factors $t_{1i}$, $(i = 1, 2, \cdots k_{12})$, and those terms containing the factors $v$ or $t_{2j}$, $(j = k_{12}+1, k_{12}+2, \cdots n_2-1)$. Further, since the order of integration here is immaterial, the equality of the $k_{12}$ integrals of the first group follows readily. Similarly, the equality of the $n_2 - k_{12}$ integrals of the second group follows. The expression (1.6) may then be written

$$(1.7) \quad \bar{x}_2 = \Big\{ k_{12} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{11} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v) dv dt_{2,n_2-1} \cdots dt_{11}$$

$$+ (n_2 - k_{12}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} v \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v) dv d_{2,n_2-1} \cdots dt_{11} \Big\}$$

$$\Big/ \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v) dv dt_{2,n_2-1} \cdots dt_{11} \Big\}$$

In (1.7), it is clear that the integrations with respect to each $t_{2j}$ may be effected immediately, making use of $\int_{-\infty}^{\infty} f(v) dv = 1$ In the first term of the numerator and in the denominator the variable $v$ may likewise be integrated out. The denominator is now equal to (1.11), the marginal distribution function of $x_1$. In the second term of the numerator, $v \cdot f(v)$ is independent of the remaining factors, and $\int_{-\infty}^{\infty} v f(v) dv$ is a constant which we shall denote by $M$. This second term of the numerator is now equal to $(n_2 - k_{12})M$ times the marginal distribution function of $x_1$

Hence, we have now reduced the expression (1.7) for $\bar{x}_2$ to the following form:

$$\bar{x}_2 = k_{12} I_{n_1} + (n_2 - k_{12}) M, \tag{1.8}$$

where
$$I_{n_1} = \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{11}\, \theta(x_1, t_{11}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{11}}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta(x_1, t_{11}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{11}}.$$

To evaluate $I_{n}$, let $t_{11} = x_1 - u - t_{12} - \cdots - t_{1,n_1-1}$.

Then
$$I_{n_1} = \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1\, \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{12}\, du}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{12}\, du}$$
$$- \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u\, \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{12}\, du}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{12}\, du}$$
$$- \sum_{j=2}^{n_1-1} \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{1j}\, \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{12}\, du}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{12}\, du}.$$

The first term in the above expression for $I_{n_1}$ is equal to $x_1$. Each of the remaining $n_1-1$ terms is equal to $I_{n_1}$. Hence

$$I_{n_1} = x_1 - (n_1 - 1) I_{n_1},$$

and

$$I_{n_1} = \frac{x_1}{n_1}.$$

From (1.8) and (1.9), we have

$$\bar{x}_2 = \frac{k_{12} x_1}{n_1} + (n_2 - k_{12}) M.$$

In exactly the same manner, we may show that

$$\bar{x}_1 = \frac{k_{12} x_2}{n_2} + (n_1 - k_{12}) M.$$

Making use of the fact that in the case of linear regression the square of the correlation coefficient is equal to the product of the slopes of the two lines of regression, we obtain

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{\frac{1}{2}}},$$

which completes the proof of the theorem.

*Corollary.* If $x$ and $y$ are each the sum of $n$ independent random values of a variable $t$ from a universe characterized by $f(t)$, and have $k$ of these values in common, the coefficient of correlation between $x$ and $y$ is equal to the ratio of the number of values of $t$ held in common to the total number composing each principal variable. Thus, $r_{xy} = \frac{k}{n}$.

This corollary of Theorem I was proved by Kapteyn[8] for the special case of a normal parent distribution of the variable $t$.

*Illustration.* As a simple illustration of the application of the foregoing theorem, let us consider the case where $x_1 = t_{11} + t_{12}$, $x_2 = t_{11} + t_{22}$ with $t_{11}$, $t_{12}$, $t_{22}$, as independent random drawings of $t$ from the Gaussian distribution,

$$f(t) = (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}}$$

From (1.11), the marginal distribution of $x_1$ is

$$G_1(x_1) = (4\pi)^{-\frac{1}{2}} e^{-\frac{x_1^2}{4}}.$$

Similarly, the marginal distribution of $x_2$ is

$$G_2(x_2) = (4\pi)^{-\frac{1}{2}} e^{-\frac{x_2^2}{4}}$$

The correlation surface, $w = F(x_1, x_2)$, obtained by applying (1.2), is

$$F(x_1, x_2) = \frac{e^{\frac{(x_1^2 - x_1 x_2 + x_2^2)}{3}}}{(2\pi\, 3^{\frac{1}{2}})},$$

a normal correlation surface with $r_{x_1 x_2} = \frac{1}{2}$.

2. The correlation among three sums. We now proceed to extend the preceding theorem to more than two sums. Let us define a third sum, or principal variable, $x_3$, as the sum of $k_{23}$

[8] J. C. Kapteyn, loc. cit.

elements taken at random from the $n_2$ values of $t$ composing $x_2$ plus the sum of $n_3 - k_{23}$ independent random values of $t$ drawn from the parent population. It is apparent, then, that the marginal distributions of $x_1$, $x_2$, and $x_3$, and the correlation surfaces $F_1'(x_1, x_2)$ and $F_2'(x_2, x_3)$ will be formed exactly as were those of $x_1$ and $x_2$ in Theorem I. From this theorem, we are at once in a position to write the equations of the lines of regression and the coefficients of correlation for these surfaces. The surface $w = F'(x_1, x_3)$ remains to be investigated, as does the four-dimensional surface, $v = \psi(x_1, x_2, x_3)$, which may be obtained in almost the same manner.

Theorem II. *Given* $f(t)$ *and* $x_1$, $x_2$, $x_3$, *as defined above. Let* $\theta_{x_1}$ *be defined as in* (1.5). *Let*

$$\phi(x, t_{11}, \cdots t_{1,k_{23}-g}, t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1}) = f(t_{2,k_{23}-g+1}) \cdots f(t_{2k_{23}}) f(t_{3,k_{23}+1}) \cdots f(t_{3,n_3-1}) \times f(x - t_{11} - \cdots - t_{1,k_{23}-g} - t_{2,k_{23}-g+1} - \cdots - t_{2k_{23}} - t_{3,k_{23}+1} - \cdots - t_{3,n_3-1}).$$

*If* $f(t)$ *is a probability function of the first kind, then the expression for the simultaneous distribution of* $x_1$ *and* $x_3$ *is*

$$F(x_1, x_3) = \sum_{g=0}^{k_{23}} \left\{ \binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \theta_{x_1}\phi(x_3, t_{11}, \cdots t_{1,k_{23}-g}, t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1})\, dt_{3,n_3-1} \cdots dt_{3,k_{23}+1} \times dt_{2k_{23}} \cdots dt_{2,k_{23}-g+1}\, dt_{1,n_1-1} \cdots dt_{11} \right\} \Big/ \binom{n_2}{k_{23}}, \tag{2.1}$$

*where by* $\binom{c}{d}$ *is understood the number of combinations of* $c$ *items taken* $d$ *at a time.*

*Proof*. Let us temporarily require that $k_{12} \geq k_{23}$. We shall show later that this restriction may be removed. The probability that $x_1$ and $x_3$ as defined contain $k_{23}-g$, $(g = 0, 1, 2, \cdots k_{23})$ elements in common is $\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g} \Big/ \binom{n_2}{k_{23}}$.

The probability of the occurrence of any given pair of values $(x_1, x_3)$, that is, the probability of a point falling into a given rectangle, $(x_1, x_1+\Delta x_1, x_3, x_3+\Delta x_3)$, is the sum of the probabilities of all of the mutually exclusive ways in which it can occur. Each of the terms in (2.1) multiplied by $dx_1 dx_3$ consists of the integral, (derived by the method of Theorem I), which is the probability, to within infinitesimals of higher order, of the occurrence of a given pair, $(x_1, x_3)$, with a specified number of values of $t$ in common, times a coefficient which is equal to the probability of the occurrence of this specified number of values of $t$ in common. Each of the terms as a whole, then, is the probability that the given $(x_1, x_3)$ will occur with a specified number of values of $t$ in common. Hence, the expression (2.1), being the sum of the probabilities of all of the mutually exclusive ways in which $x_1$ and $x_3$ can fall within the desired rectangle, is the probability that this will occur. This establishes the theorem when $k_{12} \geqq k_{23}$

If $k_{12} < k_{23}$, then the maximum number of values of $t$ which $x_1$ and $x_3$ can have in common is $k_{12}$. The expression for $F(x_1, x_3)$ in this case, then, consists of the sum of all of the terms of (2.1) beginning with the term where $x_1$ and $x_3$ have $k_{12}$ values of $t$ in common and continuing to include the term derived from the case where they have no values of $t$ in common. Equation (2.1), however, in its present form may be considered as a correct formal expression for the correlation surface even when $k_{12} < k_{23}$, since in this case all of the coefficients of the terms where $x_1$ and $x_3$ are to have more than $k_{12}$ values of $t$ in common are zero. This follows from the definition $\binom{c}{d} = 0$ if $c < d$. Thus

$$\binom{k_{12}}{k_{23}} = \binom{k_{12}}{k_{23}-1} = \cdots = \binom{k_{12}}{k_{12}+1} = 0 \text{ if } k_{12} < k_{23}.$$

Hence, we may now remove the restriction that $k_{12} \geqq k_{23}$. This establishes the theorem.

We are now in a position to write down the surface

$$v = \psi(x_1, x_2, x_3).$$

It is given by the following expression, where, by $t_{2,k_{23}-g+1}, \dots t_{2,k_{23}}$ is meant any $g$ values of the $t_{2j}$ :

$$\psi(x_1,x_2,x_3) = \sum_{g=0}^{k_{23}} \left\{ \binom{k_{12}}{k_{23}-g} \binom{n_2-k_{12}}{g} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta_{x_1} f(t_{2,k_{12}+1}) \cdots f(t_{2,n_2-1}) \right.$$

$$\times f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1}) f(t_{3,k_{23}+1}) \cdots f(t_{3,n_3-1})$$

$$\times f(x_3 - t_{11} - \cdots - t_{1,k_{23}-g} - t_{2,k_{23}-g+1} - \cdots - t_{2,k_{23}} - t_{3,k_{23}+1} - \cdots - t_{3,n_3-1})$$

$$\left. dt_{3,n_3-1} \cdots dt_{3,k_{23}+1}\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11} \right\}.$$

Theorem III. *The regression curves of $x_3$ on $x_1$ and of $x_1$ on $x_3$ for the correlation surface $w = F(x_1, x_3)$, defined in Theorem II, are linear and are given, respectively, by the following equations:*

$$(2.21) \qquad \bar{x}_3 = \frac{k_{12} k_{23} x_1}{n_1 n_2} + \frac{(n_2 n_3 - k_{12} k_{23}) M}{n_2},$$

*and*

$$(2.22) \qquad \bar{x}_1 = \frac{k_{12} k_{23} x_3}{n_2 n_3} + \frac{(n_1 n_2 - k_{12} k_{23}) M}{n_2},$$

*where $M$ is defined as in Theorem I. Further, the coefficient of correlation between $x_1$ and $x_3$ is*

$$(2.3) \qquad r_{x_1 x_3} = \frac{k_{12} k_{23}}{n_2 (n_1 n_3)^{\frac{1}{2}}} = r_{x_1 x_2}\, r_{x_2 x_3}.$$

*Proof.* As in the proof of Theorem I, we set up the expression for the locus of the ordinate of the centroid of a section of the surface for a fixed $x_1$. We have

$$\bar{x}_3 = \frac{\int_{-\infty}^{\infty} x_3 F(x_1, x_3)\, dx_3}{\int_{-\infty}^{\infty} F(x_1, x_3)\, dx_3}$$

where $F(x_1, x_3)$ is given by (2.1). From the definition of a

marginal distribution, we know that $\int_{-\infty}^{\infty} F(x_1, x_3)dx_3$ reduces to (1.11), the marginal distribution of $x_1$. Let us now write the expression for $\bar{x}_3$ as the sum of $k_{23}+1$ fractions. Thus

$$(2.4)\quad \bar{x}_3 = \sum_{g=0}^{k_{23}} \binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Big\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_3\,\theta_{x_1}$$

$$\times \phi(x_3, t_{11}, \cdots t_{1,k_{23}-g}, t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1})$$

$$\times dt_{3,n_3-1}\cdots dt_{3,k_{23}+1}\, dt_{2k_{23}}\cdots dt_{2,k_{23}-g+1}\, dt_{1,n_1-1}\cdots dt_{11}\Big\}$$

$$\Big/ \Big\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\theta_{x_1}\, dt_{1,n_1-1}\cdots dt_{11}\binom{n_2}{k_{23}}.$$

Hereafter, we shall call an expression of the form

$$\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Big/\binom{n_2}{k_{23}}$$

a "probability coefficient." Then (2.4) is the sum of products, each of which is a probability coefficient times an expression which is equivalent to the expression for $\bar{x}_3$ for the simple case where $x_3$ would be derived directly from $x_1$ by the drawing of $k_{23}-g$ values of $t$ from $x_1$. These latter expressions, by the application of Theorem I, may each be written in the same form as (1.3). Hence, (2.4) has been reduced to

$$(2.5)\quad \bar{x}_3 = x_1\Big[\frac{1}{n_1}\sum_{g=0}^{k_{23}-1}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}(k_{23}-g)\Big]\Big/\binom{n_2}{k_{23}}$$

$$+M\Big[\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}(n_3-k_{23}+g)\Big]\Big/\binom{n_2}{k_{23}}$$

$$=\frac{x_1 k_{12}}{n_1}\Big[\sum_{g=0}^{k_{23}-1}\binom{k_{12}-1}{k_{23}-g-1}\binom{n_2-k_{12}}{g}\Big]\Big/\binom{n_2}{k_{23}}$$

$$+M\left[(n_3-k_{23})\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\right.$$

$$\left.+\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}g\right]\Bigg/\binom{n_2}{k_{23}}.$$

By the use of a well-known theorem of combinatory analysis,[9] we have that

$$\frac{1}{n_1}\sum_{g=0}^{k_{23}-1}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}(k_{23}-g)\Bigg/\binom{n_2}{k_{23}}=\frac{k_{12}}{n_1}\binom{n_2-1}{k_{23}-1}\Bigg/\binom{n_2}{k_{23}}=\frac{k_{12}k_{23}}{n_1n_2},$$

and

$$(n_3-k_{23})\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Bigg/\binom{n_2}{k_{23}}=\binom{n_2}{k_{23}}(n_3-k_{23})\Bigg/\binom{n_2}{k_{23}}=(n_3-k_{23}).$$

Moreover,

$$\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}g\Bigg/\binom{n_2}{k_{23}}=(n_2-k_{12})\sum_{g=1}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}-1}{g-1}\Bigg/\binom{n_2}{k_{23}},$$

which reduces to

$$(n_2-k_{12})\binom{n_2-1}{k_{23}-1}\Bigg/\binom{n_2}{k_{23}}=(n_2-k_{12})k_{23}/n_2$$

by the same theorem of combinatory analysis.

Hence, (2.5) becomes

$$\bar{x}_3=\frac{k_{12}k_{23}x_1}{n_1n_2}+\frac{(n_2n_3-k_{12}k_{23})M}{n_2}.$$

In exactly the same manner, we may show that

$$\bar{x}_1=\frac{k_{12}k_{23}x_3}{n_2n_3}+\frac{(n_1n_2-k_{12}k_{23})M}{n_2}.$$

We then obtain the coefficient of correlation from the slopes of these lines. It is

$$r_{x_1x_3}=\frac{k_{12}k_{23}}{n_2(n_1n_3)^{1/2}}=r_{x_1x_2}\,r_{x_2x_3}.$$

This completes the proof of the theorem, since

$$r_{x_1x_2}=\frac{k_{12}}{(n_1n_2)^{1/2}}\qquad r_{x_2x_3}=\frac{k_{23}}{(n_2n_3)^{1/2}}.$$

3. The correlation among $p$ sums. We now extend our discussion to $p$ principal variables, forming each successive one

[9] E. Netto, "Lehrbuch der Combinatorik," (1901), pp. 12-13.

in the same manner in which $x_2$ and $x_3$ were formed above; that is, $x_i, (i = 2, 3 \cdots p)$, is equal to the sum of $k_{i-1, i}$ random drawings of $t$ from the constituent values of $t$ forming $x_{i-1}$, plus the sum of $n_i - k_{i-1, i}$ independent random drawings of $t$ directly from the universe characterized by $f(t)$. The correlation surface, $w = F(x_1, x_p)$, can at once be written in the same manner as the surface considered in Theorem II. That is, each term of the expression for $F(x_1, x_p)$, multiplied by $dx_1, dx_p$, consists of an iterated integral which represents the probability, to within infinitesimals of higher order, of the occurrence of a given pair, $(x_1, x_p)$, with a specified number of values of $t$ in common, times a probability coefficient which represents the probability of the occurrence of this specified number of values of $t$ in common. This same method may be employed in writing the correlation surface for any pair of principal variables. The expressions for the probability coefficients, however, become increasingly complex as the number of ways in which the two principal variables can have $0, 1, 2, \cdots$ values of $t$ in common increases.

The following theorem can be proved by mathematical induction. The proof is not difficult, though tedious, and on that account will not be presented here.

Theorem IV. *If $f(t)$ is a probability function of the first kind, and $F(x_1, x_p)$ is the simultaneous law of distribution of $x_1$ and $x_p$, then the regression of $x_1$ on $x_p$ and of $x_p$ on $x_1$ are linear and are given, respectively, by the following equations:*

$$(3.1)\quad \bar{x}_p = \frac{k_{12}\, k_{23} \cdots k_{p-1,\, p}}{n_1\, n_2 \cdots n_{p-1}}\, x_1 + \frac{n_2\, n_3 \cdots n_p - k_{12}\, k_{23} \cdots k_{p-1,\, p}}{n_2\, n_3 \cdots n_{p-1}}\, M,$$

$$(3.2)\quad \bar{x}_1 = \frac{k_{12}\, k_{23} \cdots k_{p-1,\, p}}{n_2\, n_3 \cdots n_p}\, x_p + \frac{n_1\, n_2 \cdots n_{p-1} - k_{12}\, k_{23} \cdots k_{p-1,\, p}}{n_2\, n_3 \cdots n_{p-1}}\, M.$$

Further, the coefficient of correlation between $x_1$ and $x_p$ is

$$(3.3)\qquad r_{x_1 x_p} = \frac{k_{12}\, k_{23} \cdots k_{p-1,\, p}}{n_2\, n_3 \cdots n_{p-1} (n_1 n_p)^{\frac{1}{2}}} = r_{x_1 x_2} \cdot r_{x_2 x_3} \cdots r_{x_{p-1} x_p}.$$

II. Sums of elements drawn from a universe characterized by a probability function of the second kind.

4. The correlations between two sums. Let $f(t)$, a probability function of the second kind, characterize the distribution of the variable $t$. Let the principal variable $x_1$ be defined as the sum of $n_1$ independent values of $t$ drawn at random. Further, let the principal variable $x_2$ be defined as the sum of $k_{12}$ random values of the $n_1$ values of $t$ composing $x_1$ and of $n_2 - k_{12}$ independent random values of $t$ taken directly from the universe characterized by $f(t)$.

Theorem V. *Given the sums $x_1$ and $x_2$ as defined above with $k_{12}$ random elements in common.*

a) *The marginal distributions of $x_1$ and $x_2$ are given, respectively, by*

$$(4.11)\quad G_1(x_1) = \int_0^{x_1}\int_0^{x_1 - t_{11}} \cdots \int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} f(t_{11}) \cdots f(t_{1,n_1-1})$$

$$\times f(x_1 - t_{11} - \cdots - t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{11},$$

*and*

$$(4.12)\quad G_2(x_2) = \int_0^{x_2}\int_0^{x_2 - t_{11}} \cdots \int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-2}}$$

$$\times f(t_{11}) \cdots f(t_{1k_{12}}) f(t_{2,k_{12}+1}) \cdots f(t_{2,n_2-1})$$

$$\times f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1})\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1k_{12}} \cdots dt_{11}.$$

b) *The correlation surface, $w = F(x_1, x_2)$, which is in two distinct parts joined along the plane $x_1 - x_2 = 0$, is given by*

(4.2a)

$$F_1(x_1, x_2) = \int_0^{x_2}\int_0^{x_2 - t_{11}} \cdots \int_0^{x_2 - t_{11} - \cdots - t_{1,k_{12}-1}} \int_0^{x_1 - t_{11} - \cdots - t_{1,k_{12}-1} - t_{1k_{12}}}$$

$$\int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} \int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}}} \int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1}}$$

$$\int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-2}} \theta(x_1, t_{11}, \cdots t_{1,n_1-1})$$

$$\times \phi(x_2, t_{11}, \cdots t_{1k_{12}}, t_{2,k_{12}+1}, \cdots t_{2,n_2-1})\, dt_{2,n_2-1} \cdots$$

$$\times\, dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11},$$

$$(x_2 \leqq x_1 < \infty);$$

(4.2b)

$$F_2(x_1, x_2) = \int_0^{x_1}\int_0^{x_1-t_{11}} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}}$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}} \cdots \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}}$$

$$\times\, \theta(x_1, t_{11}, \cdots t_{1,n_1-1})\, \phi(x_2, t_{11}, \cdots t_{1k_{12}}, t_{2,k_{12}+1}, \cdots t_{2,n_2-1})$$

$$\times\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11},$$

$$(x_1 \leqq x_2 < \infty).$$

c) *The regression curves of $x_2$ on $x_1$ and of $x_1$ on $x_2$ are linear and are given, respectively, by the following equations:*

(1.31) $$\bar{x}_2 = \frac{k_{12} x_1}{n_1} + (n_2 - k_{12}) M,$$

*and*

(1.32) $$\bar{x}_1 = \frac{k_{12} x_2}{n_2} + (n_1 - k_{12}) M,$$

*where* $$M = \int_0^\infty t f(t)\, dt.$$

*Hence, the cofficient of correlation between $x_1$ and $x_2$ is*

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{\frac{1}{2}}}.$$

*Proof.* The proof for the marginal distributions of $x_1$ and of $x_2$ are given by Craig[10] and need not be repeated here. The expressions for the correlation surface are derived by a simple extension of the same method to two independent variables. The

[10] Allen T. Craig, loc. cit., p. 356.

limits of integration may be easily verified.

As in the proof of Theorem I, the regression of $x_2$ on $x_1$ is given by the locus of the ordinate of the centroid of the section of the surface for a given $x_1$. However, as the surface here is in two distinct, but connected, parts, we have two terms in both numerator and denominator. The expression for $\bar{x}_2$ is

$$(4.3) \qquad \bar{x}_2 = \frac{\int_0^{x_1} x_2 F_1(x_1, x_2)\,dx_2 + \int_{x_1}^{\infty} x_2 F_2(x_1, x_2)\,dx_2}{\int_0^{x_1} F_1(x_1, x_2)\,dx_2 + \int_{x_1}^{\infty} F_2(x_1, x_2)\,dx_2},$$

where $F_1(x_1, x_2)$ and $F_2(x_1, x_2)$ are defined by (4.2a) and (4.2b), respectively.

In the paragraphs immediately following, we shall be concerned principally with interchanging the order of integration, with the accompanying changes in the limits. It will be convenient to write the differential immediately following its respective integral sign. Consider the first term of the numerator. Successive interchanging the order of integration between integration with respect to $x_2$ and with respect to $t_{11}, t_{12}, \ldots t_{1k_{12}}$ respectively, and making the appropriate changes in the limits, we get, writing $\Phi_{x_2}$ for $\Phi(x_2, t_{11}, \cdots t_{1k_{12}}, t_{2,k_{12}+1}, \cdots t_{2,n_2-1})$

$$(4.4) \quad \int_0^{x_1} dt_{11} \int_0^{x_1-t_{11}} dt_{12} \cdots \int_0^{x_1-t_{11}-t_{12}-\cdots-t_{1,k_{12}-1}} dt_{1k_{12}} \int_{t_{11}+t_{12}+\cdots+t_{1k_{12}}}^{x_1} dx_2$$

$$\int_0^{x_1-t_{11}-\cdots-t_{1k_{12}}} dt_{1,k_{12}+1} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1} \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}} dt_{2,k_{12}+1}$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}} dt_{2,k_{12}+2} \cdots \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}} dt_{2,n_2-1}$$

$$\times\, x_2\, \theta_{x_1}\, \Phi_{x_2}.$$

Now consider the second term of the numerator of (4.3). As he limits are constants with respect to the variables of integration

$x_2, t_{11}, \ldots t_{1K_{12}}$, we may interchange the order of integration successively until we have

(4.5)

$$\int_0^{x_1} dt_{11} \int_0^{x_1-t_{11}} dt_{12} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,K_{12}-1}} dt_{1K_{12}} \int_{x_1}^{\infty} dx_2$$

$$\int_0^{x_1-t_{11}-\cdots-t_{1K_{12}}} dt_{1,K_{12}+1} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1K_{12}}} dt_{2,K_{12}+1} \int_0^{x_2-t_{11}-\cdots-t_{1K_{12}}-t_{2,K_{12}+1}} dt_{2,K_{12}+2} \cdots$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1K_{12}}-t_{2,K_{12}+1}-\cdots-t_{2,n_2-2}} dt_{2,n_2-1}\, x_2\, \theta_{x_1}\, \phi_{x_2}\,.$$

We may now combine the first and second terms, (4.4) and (4.5), getting

$$\int_0^{x_1} dt_{11} \int_0^{x_1-t_{11}} dt_{12} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,K_{12}-1}} dt_{1K_{12}} \int_{t_{11}+t_{12}+\cdots+t_{1K_{12}}}^{\infty} dx_2$$

$$\int_0^{x_1-t_{11}-\cdots-t_{1K_{12}}} dt_{1,K_{12}+1} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1K_{12}}} dt_{2,K_{12}+1} \int_0^{x_2-t_{11}-\cdots-t_{1K_{12}}-t_{2,K_{12}+1}} dt_{2,K_{12}+2} \cdots$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1K_{12}}-t_{2,K_{12}+1}-\cdots-t_{2,n_2-2}} dt_{2,n_2-1}\, x_2\, \theta_{x_1}\, \phi_{x_2}\,.$$

As the limits of integration are constant with respect to the variables $x_2$ and $t_{1,K_{12}+1}, \ldots t_{1,n_1-1}$, we may at once interchange successively the orders of integration with respect to $x_2$ and with respect to $t_{2,K_{12}+1}, t_{2,K_{12}+2}, \ldots \ldots t_{2,n_2-1}$, respectively, making the proper changes in the limits. We then have

(4.6)

$$\int_0^{x_1} dt_{11} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1} \int_0^{\infty} dt_{2,K_{12}+1} \cdots$$

$$\int_0^{\infty} dt_{2,n_2-1} \int_{t_{11}+\cdots+t_{1K_{12}}+t_{2,K_{12}+1}+\cdots+t_{2,n_2-1}}^{\infty} dx_2\, x_2\, \theta_{x_1}\, \phi_{x_2}\,.$$

The denominator of (4.3) may be reduced to this same form except for the absence of the factor $x_2$ in the integrand.

Let us make the transformation

$$v = x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1},$$

as was done in the proof of Theorem I. The limits $t_{11}+\cdots+t_{2,n_2-1}$ to $\infty$ on $x_2$ now become $0$ to $\infty$ on $v$. We have now reduced (4.3) to the following form:

(4.7)

$$\bar{x}_2 = \Big\{\Big[\int_0^{x_1}\cdots\int_0^{\infty} t_{11} + \int_0^{x_1}\cdots\int_0^{\infty} t_{12} + \cdots\cdots \int_0^{x_1}\cdots\int_0^{\infty} t_{1k_{12}} + \int_0^{x_1}\cdots\int_0^{\infty} t_{2,k_{12}+1} + \cdots + \int_0^{x_1}\cdots\int_0^{\infty} t_{2,n_2-1} + \int_0^{x_1}\cdots\int_0^{\infty} v\Big]\theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dv\, dt_{2,n_2-1}\cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1}\cdots dt_{11}\Big\} \Big/ \Big\{\int_0^{x_1}\cdots\int_0^{\infty}\theta_{x_1}\prod_{j=k_{12}+1}^{n_2-1} f(t_{2j})f(v)\,dv\,dt_{2,n_2-1}\cdots dt_{2,k_{12}+1}\,dt_{1,n_1-1}\cdots dt_{11}\Big\}.$$

The denominator reduces at once to $G(x_1)$ in (4.11). As in the proof of Theorem I directly following equation (1.6), it will be noted that the terms of the numerator fall into two groups: those $k_{12}$ terms containing the factor $t_{1i}$, $(i = 1, 2, \ldots\ldots k_{12})$, and the $n_2 - k_{12}$ terms containing the factor $v$ or $t_{2j}$, $(j = k_{12}+1, \ldots\ldots n_2-1)$. As the limits of integration with respect to each of these letter variables are $0$ and $\infty$, and since complete interchangeability of the order of integration is then permissible, it is readily seen that any two of these $n_2 - k_{12}$ terms are equivalent. The sum of the entire group, then, may be written

(4.8)

$$(n_2 - k_{12})\int_0^{x_1} dt_{11}\cdots\int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} dt_{1,n_1-1}\int_0^{\infty} dt_{2,k_{12}+1}\cdots \int_0^{\infty} dt_{2,n_2-1}\int_0^{\infty} dv\; v\,\theta_{x_1}\prod_{j=k_{12}+1}^{n_2-1} f(t_{2j})\, f(v).$$

In (4.8), it is clear that the integrations with respect to each $t_{2j}$ may be effected immediately by making use of the hypothesis that $\int_0^\infty f(t)\,dt = 1$. This leaves $v f(v)\,\theta_{x_1}$ remaining as the integrand. The $\int_0^\infty v f(v)\,dv$ is a constant which we shall designate by $M$. Removing this constant from under the integral signs leaves us merely the expression for the marginal distribution of $x_1$ times $M(n_2-k_{12})$. We then have

$$(4.9)\quad \bar{x}_2 = (n_2-k_{12})M + \sum_{i=1}^{k_{12}} \frac{\int_0^{x_1} dt_{11} \cdot \int_0^\infty dv\, t_{1i}\, \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)}{\int_0^{x_1} dt_{11} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}\, \theta_{x_1}}.$$

That each term in the summation in the right member of (4.9) is equal to any other term in the summation, follows from the complete interchangeability of the order of integration of any two consecutive variables, provided a corresponding interchange between these two variables is likewise carried out in the limits of integration. By successive interchanges of variables we may put the original $t_{11}$, $t_{12}$, .... $t_{1k_{12}}$ in any order we choose. Hence, the sum of the last $k_{12}$ terms of (4.9) may be written as $k_{12}$ times any one of them. For definiteness, select the one containing the factor $t_{11}$ in the integrand of the numerator. We may now integrate out all of the $t_{2j}$, $(j = k_{12}+1, \dots n_2-1)$, and the $v$ exactly as before. Equation (4.9) then becomes

$$\bar{x}_2 = (n_2-k_{12})M + k_{12}\frac{\int_0^{x_1} dt_{12} \cdots \int_0^{x_1-t_{12}-\cdots-t_{1,n_1-1}} dt_{11}\, t_{11}\, \theta_{x_1}}{\int_0^{x_1} dt_{12} \cdots \int_0^{x_1-t_{12}-\cdots-t_{1,n_1-1}} dt_{11}\, \theta_{x_1}},$$

or $\bar{x}_2 = (n_2-k_{12})M + k_{12} I_{n_1}$

It is not difficult to show that $I_{n_1} = \frac{x_1}{n_1}$ Hence, we have

$$\bar{x}_2 = \frac{k_{12}\, x_1}{n_1} + (n_2-k_{12})M.$$

In exactly the same manner, we may show that

$$\bar{x}_1 = \frac{k_{12}\, x_2}{n_2} + (n_1-k_{12})M.$$

The coefficient of correlation between $x_1$ and $x_2$ is

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{1/2}},$$

which completes the proof of the theorem.

*Illustration.* Consider the two sums, $x_1 = t_{11} + t_{12}$, and $x_2 = t_{11} + t_{22}$, with $t_{11}$, $t_{12}$, $t_{22}$, as random drawings of $t$ from the distribution characterized by the function $f(t) = e^{-t}$ for $t$ on the range $0$ to $\infty$. From (4.11), the marginal distribution of $x_1$ is

$$G_1(x_1) = x_1 e^{-x_1}.$$

Similarly, the marginal distribution of $x_2$ is

$$G_2(x_2) = x_2 e^{-x_2}.$$

The correlation surface, obtained by applying (4.2a) and (4.2b), is

$$F_1(x_1, x_2) = e^{-x_1}(1 - e^{-x_2}), \quad (0 \leq x_2 \leq x_1);$$

and

$$F_2(x_1, x_2) = e^{-x_2}(1 - e^{-x_1}), \quad (x_1 \leq x_2 < \infty).$$

5. The correlation among more than two sums. We shall state, without proof, the following theorems.

Theorem VI. *Given a probability function, $f(t)$, of the second kind, and three principal variables, $x_1$, $x_2$, $x_3$, defined as for Theorem II. Then the correlation surface $w = F(x_1, x_3)$ is given by*

(5.1a)

$$F_1(x_1, x_3) = \frac{1}{\binom{n_2}{k_{23}}} \sum_{g=0}^{k_{23}} \binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g} \int_0^{x_3} dt_{11} \cdots \int_0^{x_3 - t_{11} - \cdots - t_{1,k_{23}-g-1}} dt_{1,k_{23}-g}$$

$$\int_0^{x_1 - t_{11} - \cdots - t_{1,k_{23}-g}} dt_{1,k_{23}-g+1} \cdots \int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} dt_{1,n_1-1}$$

$$\int_0^{x_3 - t_{11} - \cdots - t_{1,k_{23}-g}} dt_{2,k_{23}-g+1} \cdots$$

$$\int_0^{x_3 - t_{11} - \cdots - t_{1,k_{23}-g} - t_{2,k_{23}-g+1} - \cdots - t_{2,k_{23}-1}} dt_{2k_{23}}$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2k_{23}}} dt_{3,k_{23}+1}\cdots$$

$$\int_0^{x_3-t_{11}-\cdots-t_{3,n_3-2}} dt_{3,n_3-1}\ \theta_{x_1}\ \Phi(x_3, t_{11}, \cdots t_{1,k_{23}-g},$$

$$t_{2,k_{23}-g+1}, \cdots t_{2k_{23}},\ t_{3,k_{23}+1}, \cdots t_{3,n_3-1}),$$

$$(x_3 \leqq x_1 < \infty);$$

*and*

(5.1b)

$$F_2(x_1,x_3)=\frac{1}{\binom{n_2}{k_{23}}}\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\int_0^{x_1} dt_{11}\cdots\int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}} dt_{2,k_{23}-g+1}\cdots$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2,k_{23}-1}} dt_{2k_{23}}$$

$$\int_0^{x_3-t_{11}-\cdots-t_{2,k_{23}-1}-t_{2,k_{23}}} dt_{3,k_{23}+1}\cdots$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2k_{23}}-t_{3,k_{23}+1}-\cdots-t_{3,n_3-2}} dt_{3,n_3-1}$$

$$\theta_{x_1}\Phi(x_3, t_{11}, \cdots t_{1,k_{23}-g}, t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1}),$$

$$(x_1 \leqq x_3 < \infty).$$

Theorem VII. *The regression curves of $x_3$ on $x_1$ and of $x_1$ on $x_3$ of the correlation surface in Theorem VI are linear and are given, respectively, by the following equations:*

(2.21)
$$\bar{x}_3 = \frac{k_{12}k_{23}x_1}{n_1n_2} + \frac{(n_2n_3-k_{12}k_{23})M}{n_2},$$

*and*

(2.22)
$$\bar{x}_1 = \frac{k_{12}k_{23}x_3}{n_2n_3} + \frac{(n_1n_2-k_{12}k_{23})M}{n_2},$$

*where $M$ is defined as in Theorem V. Further, the coefficient of correlation between $x_1$ and $x_3$ is*

$$(2.3)\qquad r_{x_1x_3} = \frac{k_{12}k_{23}}{n_2(n_1n_3)^{1/2}} = r_{x_1x_2}\,r_{x_2x_3}\,.$$

Theorem VIII. The statement of this theorem differs from that of Theorem IV only in that $f(t)$ is now to be a probability function of the second kind.

III. Sums of elements drawn from a universe characterized by a probability function of the third kind.

6. The correlation between two sums. We shall now consider principal variables defined as the sums of values of $t$ drawn from a universe characterized by $f(t)$, a probability function of the third kind, defined on the range $0$ to $a$, and with

$$\int_0^a f(t)\,dt = 1.$$

The correlation surfaces are not developed with the same degree of generality as were those in the preceding pages because of the tediousness of the labor involved and the complexity of the correlation surface, which may consist of many sections joined together. Thus, if $x$ is the sum of $m$ values of $t$ and $y$ the sum of $n$, all drawn from a universe characterized by a probability function of the third kind, the correlation surface, $w = F(x,y)$, consists of $2(mn-1)$ sections, each having its own equation. Hence, only the case where $x$ and $y$ each consist of the sum of two values of $t$, with one of these held in common, will be considered here.

Theorem IX. *Let $f(t)$, a probability function of the third kind, characterize the distribution of a variable $t$. Let the principal variables $x$ and $y$ be defined by the relations $x = t_{11} + t_{12}$, $y = t_{11} + t_{22}$, where $t_{11}$, $t_{12}$, $t_{22}$, are independent random drawings of $t$ from the universe.*

a.) *The marginal distributions of $x$ and of $y$ are given by*

(6.11)
$$G_1(x) = \int_0^x f(t)f(x-t)dt, \qquad (0 \leqq x \leqq a);$$
$$= \int_{x-a}^a f(t)f(x-t)dt, \qquad (a \leqq x \leqq 2a);$$

*and*

(6.12)
$$G_2(y) = \int_0^y f(t)f(y-t)dt, \qquad (0 \leqq y \leqq a);$$
$$= \int_{y-a}^a f(t)f(y-t)dt, \qquad (a \leqq y \leqq 2a).$$

b) *The correlation surface,* $w = F(x,y)$ *, is given by*

(6.2)
$$F(x,y) = \int_0^y f(t)f(x-t)f(y-t)dt, \quad (0 \leqq y \leqq x \leqq a);$$
$$= \int_0^x f(t)f(x-t)f(y-t)dt, \quad (0 \leqq x \leqq y \leqq a);$$
$$= \int_{y-a}^x f(t)f(x-t)f(y-t)dt, \quad (a \leqq y \leqq x+a \leqq 2a);$$
$$= \int_{x-a}^y f(t)f(x-t)f(y-t)dt, \quad (0 \leqq x-a \leqq y \leqq a);$$
$$= \int_{x-a}^a f(t)f(x-t)f(y-t)dt, \quad (a \leqq y \leqq x \leqq 2a);$$
$$= \int_{y-a}^a f(t)f(x-t)f(y-t)dt, \quad (a \leqq x \leqq y \leqq 2a).$$

In a) and b) above, the subscripts have been omitted from the $t_{ii}$

c) *The regression curves of* $y$ *on* $x$ *and of* $x$ *on* $y$ *are linear and are given, respectively, by the following equations:*

(6.31)
$$\bar{y} = \frac{x}{2} + M,$$

(6.32) *and*
$$\bar{x} = \frac{y}{2} + M,$$

*where*
$$M = \int_0^a t\, f(t)dt.$$

*Hence, the coefficient of correlation between* $x$ *and* $y$ *is* $\frac{1}{2}$ .

This theorem is a direct generalization of Rietz's paper in Biometrika cited in the introduction to this paper. The proof may be supplied by the reader.

*Illustration.* Let us consider the rectangular distribution given by $f(t) = \frac{1}{a}$, for $t$ on the range $0$ to $a$, and a to $0$. This is the parent distribution in Rietz's case when $a = 1$. From (6.11), the marginal distribution of $x$ is

$$\begin{aligned} G_1(x) &= \frac{x}{a^2}, && (0 \leqq x \leqq a); \\ &= \frac{(2a-x)}{a^2}, && (a \leqq x \leqq 2a). \end{aligned}$$

Similarly, the marginal distribution of $y$ is

$$\begin{aligned} G_2(y) &= \frac{y}{a^2}, && (0 \leqq y \leqq a); \\ &= \frac{(2a-y)}{a^2}, && (a \leqq y \leqq 2a). \end{aligned}$$

The application of (6.2) yields

$$\begin{aligned} F(x,y) &= \frac{y}{a^3}, && (0 \leqq y \leqq x \leqq a); \\ &= \frac{x}{a^3}, && (0 \leqq x \leqq y \leqq a); \\ &= \frac{(x-y+a)}{a^3}, && (a \leqq y \leqq x+a \leqq 2a); \\ &= \frac{(y-x+a)}{a^3}, && (0 \leqq x-a \leqq y \leqq a); \\ &= \frac{(2a-x)}{a^3}, && (a \leqq y \leqq x \leqq 2a); \\ &= \frac{(2a-y)}{a^3}, && (a \leqq x \leqq y \leqq 2a). \end{aligned}$$

These results, obtained directly by the use of Theorem IX, agree with those obtained by Rietz in the above-mentioned paper.

Carl H. Fischer.

# ON THE CORRELATION BETWEEN CERTAIN AVERAGES FROM SMALL SAMPLES*

*By*

ALLEN T. CRAIG

1. *Introduction.* It is well known that no correlation exists between the arithmetic mean and standard deviation of samples drawn at random from a normal universe. However, there seems to be in the literature no treatment of the correlation between other averages either for normal or non-normal universes. In the present paper, a few simple theorems are established which make possible the determination of the type of regression of the median on the arithmetic mean, of the range on the median, and of the range on the arithmetic mean. In case the regression is linear, the coefficient of correlation may be computed.

We shall understand a probability function $f(x)$ of a real variable $x$ to be, for all values of $x$ on a range of $R$ a single-valued, non-negative, continuous function with $\int_R f(x)\,dx = 1$. Then $\int_a^b f(x)dx$ is the probability that a value of $x$ chosen at random lies in the interval $(a, b)$ where $a$ and $b$ are in $R$ and $a < b$ ; and $f(x)\,dx$ is, to within infinitesimals of higher order, the probability that a value of $x$ chosen at random lies in the interval $(x,\ x+dx)$. It will prove convenient to classify probability functions according as $R$ is the range $(-\infty, \infty)$, $(0, \infty)$, or $(0, k)$, $k>0$. In accord with this classification,[1] we shall refer to probability functions as of the first, second, and third kinds respectively. In a similar manner, we define a probability function $F(x,y)$ of two independent variables.

*Presented to the American Mathematical Society, Dec. 28, 1931.

[1]Cf. L. Bachelier, Calcul des Probabilités, p. 155.

2. The correlation between the arithmetic mean $\bar{x}$ and the range $W$.

Theorem I. *Let* $f(x)$ *be the probability function of the variable* $x$. *Let* $F_1(\bar{x}, W)$ *be that of the arithmetic mean* $\bar{x}$ *and the range* $W$ *in samples of three independent values of* $x$. *If* $f(x)$ *is a probability function of the first kind, then*

$$F_1(\bar{x}, W) = 18\int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1)\, f(x_1 - W)\, f(3\bar{x} - 2x_1 + W)\, dx_1 .$$

Proof. Let $x_1$, $x_2$, $x_3$, be the three observed values of $x$. Write

$$\begin{aligned} x_1 + x_2 + x_3 &= 3\bar{x}, \\ x_1 \qquad - x_3 &= W, \\ x_3 \le x_2 &\le x_1 \end{aligned}$$

For $\bar{x}$ assigned, $-\infty < \bar{x} < \infty$, and $W$ assigned, $0 \le W < \infty$ we must have

$$\begin{aligned} \bar{x} + \frac{W}{3} \le x_1 &\le \bar{x} + \frac{2W}{3}, \\ x_3 &= x_1 - W, \\ x_2 &= 3\bar{x} - x_1 - x_3 . \end{aligned}$$

If we consider all possible arrangements of $x_1$, $x_2$, $x_3$, we have

$$F_1(\bar{x}, W)\, d\bar{x}\, dW = 6\int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1)\, f(x_2)\, f(x_3)\, dx_1\, dx_2\, dx_3 .$$

Let

$$\begin{aligned} x_1 &= x_1 \\ x_2 &= 3\bar{x} - x_1 - x_3 , \\ x_3 &= x_1 - W. \end{aligned}$$

The absolute value of the Jacobin is $3$. Hence the theorem.

In the case of samples of four independent items $x_1$, $x_2$, $x_3$, $x_4$, the probability function $F_1(\bar{x},W)$ is given by

$$F_1(\bar{x},W)=48\int_{\bar{x}+\frac{W}{4}}^{\bar{x}+\frac{W}{2}}\int_{4\bar{x}-3x_1+W}^{x_1} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)\,dx_2\,dx_1$$

$$+48\int_{\bar{x}+\frac{W}{2}}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)\,dx_2\,dx_1.$$

We note that the probability function is made up of the sum of two parts depending on whether $x_1$ is in the interval $(\bar{x}+\frac{W}{4}, \bar{x}+\frac{W}{2})$ or in the interval $(\bar{x}+\frac{W}{2}, \bar{x}+\frac{3W}{4})$. Moreover, it may be of interest to note the overlapping of the ranges of integration of $x_2$. To prove that $F_1(\bar{x},W)$ is given as stated, we take

$$\text{(1)}\qquad \begin{aligned} &x_1+x_2+x_3+x_4=4\bar{x},\\ &x_4\leq x_3,\ x_2\leq x_1,\\ &x_1-x_4=W. \end{aligned}$$

From (1) it readily follows that

$$\text{(2)}\qquad 2x_1+x_2+x_3=4\bar{x}+W.$$

For assigned values of $\bar{x}$ and $W$, the upper limit on $x_1$ is found from (2) by taking $x_2=x_3=x_4=x_1-W$. Thus $x_1=\bar{x}+\frac{3W}{4}$. Similarly, the lower limit on $x_1$ is found from (2) by taking $x_2=x_3=x_1$. Thus $x_1=\bar{x}+\frac{W}{4}$. But $x_2$ may not always be as large as $x_1$ for all values of $x_1$. This may be seen by taking $x_2=x_1$ and $x_3=x_4=x_1-W$ in (2). This leads to $x_1=\bar{x}+\frac{W}{2}$. Thus, for $\bar{x}+\frac{W}{4}\leq x_1\leq\bar{x}+\frac{W}{2}$, we see that $x_1$ is the upper limit on $x_2$. To determine the lower limit on $x_2$ for this region of variation of $x_1$, we select $x_2$ as near $x_4=x_1-W$ as is possible without causing $x_3$ to exceed $x_1$. But $x_3=4\bar{x}-2x_1-x_2+W$. At most, then $4\bar{x}-2x_1-x_2+W=x_1$ or $x_2=4\bar{x}-3\bar{x}_1+W$. Thus we have established the limits of integration used in the first part of the sum of which $F_1(\bar{x},W)$ consists. A similar argument shows if $\bar{x}+\frac{W}{2}\leq x_1\leq\bar{x}+\frac{3W}{4}$, that

$$x_1-W\leq x_2\leq 4\bar{x}-3x_1+2W.$$

If $f(x)$ is a probability function of the second kind, we observe in samples of three independent items $x_1, x_2, x_3$, for $\bar{x}$ assigned, that $0 \le W \le 3\bar{x}$. If $0 \le W \le 3\bar{x}/2$, we have

$$\bar{x} + \frac{W}{3} \le x_1 \le \bar{x} + \frac{2W}{3},$$
$$x_2 = 3\bar{x} - 2x_1 + W,$$
$$x_3 = x_1 - W,$$

and if $\frac{3\bar{x}}{2} \le W \le 3\bar{x}$, we have

$$W \le x_1 \le \bar{x} + \frac{2W}{3},$$
$$x_2 = 3\bar{x} - 2x_1 + W,$$
$$x_3 = x_1 - W.$$

Accordingly,

$$F_1(\bar{x}, W) = 18\int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1)f(x_1-W)f(3\bar{x}-2x_1+W)dx_1, \quad 0 \le W \le \frac{3\bar{x}}{2},$$
$$= 18\int_{W}^{\bar{x}+\frac{2W}{3}} f(x_1)f(x_1-W)f(3\bar{x}-2x_1+W)dx_1, \quad \frac{3\bar{x}}{2} \le W \le 3\bar{x}:$$

In samples of four independent items $x_1, x_2, x_3, x_4$, drawn from a universe characterized by a law of probability of this kind, we find

$$F_1(\bar{x}, W) = 48\int_{\bar{x}+\frac{W}{4}}^{\bar{x}+\frac{W}{2}}\int_{4\bar{x}-3x_1+W}^{x_1} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)dx_2dx_1$$
$$+48\int_{\bar{x}+\frac{W}{2}}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)dx_2\,dx_1,$$
$$0 \le W \le \frac{4\bar{x}}{3},$$
$$=48\int_{W}^{\bar{x}+\frac{W}{2}}\int_{4\bar{x}-3x_1+W}^{x_1} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)dx_2\,dx_1$$
$$+48\int_{\bar{x}+\frac{W}{2}}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)\,dx_2\,dx_1,$$
$$\frac{4\bar{x}}{3} \le W \le 2\bar{x},$$
$$=48\int_{W}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)dx_2\,dx_1.$$
$$2\bar{x} \le W \le 4\bar{x}.$$

Finally, consider $f(x)$ to be a probability function of the third kind. In samples of three independent items $x_1, x_2, x_3$, for $0 \le \bar{x} \le k/3$, we obtain $0 \le W \le 3\bar{x}$; for $k/3 \le \bar{x} \le 2k/3$, we obtain $0 \le W \le k$; for $2k/3 \le \bar{x} \le k$, we obtain $0 \le W \le 3(k-\bar{x})$. It is fairly easy to see that for $\bar{x}$ and $W$ assigned as indicated, the following regions of selection of $x_1$ are valid:

for $0 \le \bar{x} \le k/2$ and $0 \le W \le 3\bar{x}/2$,
or for $k/2 \le \bar{x} \le k$ and $0 \le W \le 3(k-\bar{x})/2$, then
$\bar{x} + W/3 \le x_1 \le \bar{x} + 2W/3$;
for $0 \le \bar{x} \le k/3$ and $3\bar{x}/2 \le W \le 3\bar{x}$,
or for $k/3 \le \bar{x} \le k/2$ and $3\bar{x}/2 \le W \le 3(k-\bar{x})/2$, then
$W \le x_1 \le \bar{x} + 2W/3$;
for $2k/3 \le \bar{x} \le k$ and $3(k-\bar{x})/2 \le W \le 3(k-\bar{x})$
or for $k/2 \le \bar{x} \le 2k/3$ and $3(k-\bar{x})/2 \le W \le 3\bar{x}/2$, then
$\bar{x} + W/3 \le x_1 \le k$;
for $k/3 \le \bar{x} \le k/2$ and $3(k-\bar{x})/2 \le W \le k$,
or for $k/2 \le \bar{x} \le 2k/3$ and $3\bar{x}/2 \le W \le k$, then $W \le x_1 \le k$.

Thus,

$$F_1(\bar{x}, W) = 18\int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1) f(x_1 - W) f(3\bar{x} - 2x_1 + W)\,dx_1,$$

$$= 18\int_{W}^{\bar{x}+\frac{2W}{3}} f(x_1) f(x_1 - W) f(3\bar{x} - 2x_1 + W)\,dx_1,$$

$$= 18\int_{\bar{x}+\frac{W}{3}}^{k} f(x_1) f(x_1 - W) f(3\bar{x} - 2x_1 + W)\,dx_1,$$

$$= 18\int_{W}^{k} f(x_1) f(x_1 - W) f(3\bar{x} - 2x_1 + W)\,dx_1,$$

over those regions of the $\bar{x}W$-plane indicated above.

In case of samples of four independent items $x_1, x_2, x_3, x_4$, drawn from a universe characterized by a probability function of the third kind,

for $k/4 \le \bar{x} \le 3k/4$, we obtain $0 \le W \le k$; for $3k/4 \le \bar{x} \le k$, we obtain $0 \le W \le 4(k-\bar{x})$. Let us denote as follows the regions of the $\bar{x}W$-plane bounded by the given lines:

$$(A)\begin{cases} \bar{x} = 0 \\ W = \frac{4\bar{x}}{3} \\ W = \frac{4(k-\bar{x})}{3} \end{cases} \qquad (E)\begin{cases} W = \frac{4\bar{x}}{3} \\ W = 2(k-\bar{x}) \\ W = 4(k-\bar{x}) \end{cases}$$

$$(B)\begin{cases} W = \frac{4(k-\bar{x})}{3} \\ W = 2\bar{x} \\ W = \frac{4\bar{x}}{3} \end{cases} \qquad (F)\begin{cases} W = \frac{4\bar{x}}{3} \\ W = 2\bar{x} \\ W = \frac{4(k-\bar{x})}{3} \\ W = 2(k-\bar{x}) \end{cases}$$

$$(C)\begin{cases} W = 2\bar{x} \\ W = 4\bar{x} \\ W = \frac{4(k-\bar{x})}{3} \end{cases} \qquad (G)\begin{cases} W = k \\ W = 2\bar{x} \\ W = \frac{4(k-\bar{x})}{3} \end{cases}$$

$$(D)\begin{cases} W = \frac{4\bar{x}}{3} \\ W = 2(k-\bar{x}) \\ W = \frac{4(k-\bar{x})}{3} \end{cases} \qquad (H)\begin{cases} W = k \\ W = \frac{4\bar{x}}{3} \\ W = 2(k-\bar{x}) \end{cases}$$

Further, let

$$\theta = f(x_1)f(x_2)f(x_1-W)f(4\bar{x}-2x_1-x_2+W),$$

and let

$$\int_a^b\int_c^d \theta\, dx_2\, dx_1 = \begin{pmatrix} b & d \\ a & c \end{pmatrix}\theta$$

It is then not difficult to verify that

$$F_1(\bar{x},W) = 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ \bar{x}+\frac{W}{4} & 4\bar{x}-3x_1+W \end{pmatrix}\theta + \begin{pmatrix} \bar{x}+\frac{3W}{4} & 4\bar{x}-3x_1+2W \\ \bar{x}+\frac{W}{2} & x_1-W \end{pmatrix}\theta\right], \text{(A)}$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ W & 4\bar{x}-3x_1+W \end{pmatrix}\theta + \begin{pmatrix} \bar{x}+\frac{3W}{4} & 4\bar{x}-3x_1+2W \\ \bar{x}+\frac{W}{2} & x_1-W \end{pmatrix}\theta\right], \text{(B)}$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{3W}{4} & 4\bar{x}-3x_1+W \\ W & x_1-W \end{pmatrix}\theta\right], \qquad \text{(C)}$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ \bar{x}+\frac{W}{4} & 4\bar{x}-3x_1+W \end{pmatrix}\theta + \begin{pmatrix} k & 4\bar{x}-3x_1+2W \\ \bar{x}+\frac{W}{2} & x_1-W \end{pmatrix}\right], \text{(D)}$$

$$= 48\left[\begin{pmatrix} k & x_1 \\ \bar{x}+\frac{W}{4} & 4\bar{x}-3x_1+W \end{pmatrix}\right]\theta\ , \tag{E}$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ W & 4\bar{x}-3x_1+W \end{pmatrix}\right]\theta\ , \tag{F}$$

$$= 48\left[\begin{pmatrix} k & 4\bar{x}-3x_1+2W \\ W & x_1-W \end{pmatrix}\right]\theta\ , \tag{G}$$

$$= 48\left[\begin{pmatrix} k & x_1 \\ W & 4\bar{x}-3x_1+W \end{pmatrix}\right]\theta\ . \tag{H}$$

As illustrations of these theorems, let us find the correlation between the range and the mean for universes of specified types.

Example 1. Let $f(x) = e^{-x}$ $0 \leq x < \infty$.
For samples of three items, we have

$$F_1(\bar{x}, W) = 6We^{-3\bar{x}}, \quad 0 \leq W \leq \frac{3\bar{x}}{2},$$
$$= 18\left(\bar{x} - \frac{W}{3}\right)e^{-3\bar{x}}, \quad \frac{3\bar{x}}{2} \leq W \leq 3\bar{x}.$$

The distributions of the marginal totals of $W$ and $\bar{x}$ are obtained by integrating $F_1(\bar{x}, W)$ with regard to $\bar{x}$ and $W$ respectively. We readily find

$$\phi(\bar{x}) = \frac{27\bar{x}^2}{2}e^{-3\bar{x}}, \qquad 0 \leq \bar{x} < \infty,$$

and

$$\psi(W) = 2e^{-2W}(e^{W}-1), \qquad 0 \leq W < \infty,$$

as previously given by the writer.[2] For $\bar{x}$ assigned, the mean of the array of $W$ is $\overline{W}_{\bar{x}} = \frac{3\bar{x}}{2}$ . Thus the regression of $W$ on $\bar{x}$ is linear and $r = \frac{\sqrt{15}}{5}$ .

[2]American Journal of Mathematics, Vol. 54 (1932), pp. 359, 366.

Example 2. Let $f(x) = 1/k, \quad 0 \le x \le k.$

For samples of three items, we have

$$F_1(\bar{x}, W) = \frac{6W}{k^3},$$

$$= \frac{18}{k^3}\left(\bar{x} - \frac{W}{3}\right),$$

$$= \frac{18}{k^3}\left(k - \bar{x} - \frac{W}{3}\right),$$

$$= \frac{18}{k^3}(k - W)$$

over those regions of the $\bar{x}W$-plane indicated above. The marginal totals[8] are distributed in accord with

$$\varphi(\bar{x}) = \frac{27\bar{x}^2}{2k^3}, \qquad 0 \le \bar{x} \le \frac{k}{3},$$

$$= \frac{9}{2k^3}\left[-6\bar{x}^2 + 6k\bar{x} - k^2\right], \quad \frac{k}{3} \le \bar{x} \le \frac{2k}{3},$$

$$= \frac{27}{2k^3}(k - \bar{x})^2, \quad \frac{2k}{3} \le \bar{x} \le k,$$

and
$$\psi(W) = \frac{6W}{k^3}(k - W), \quad 0 \le W \le k.$$

We readily find

$$\overline{W}_{\bar{x}} = \frac{3\bar{x}}{2}, \quad 0 \le \bar{x} \le \frac{k}{3},$$

[8] Cf. H. L. Rietz, On a Certain Law of Probability of Laplace, Proc. Int. Math. Congress, Toronto (1924), pp. 795-799.

J. O. Irwin, On the Frequency Distributions of Means, etc., Biometrika, Vol. 19 (1927), pp. 225-239.

P. Hall, The Distribution of Means for Samples of Size N, Biometrika, Vol. 19 (1927), pp. 240-245.

J. Neyman and E. S. Pearson, On the Use and Distribution of Certain Test Criteria, Biometrika, Vol. 20 (1928), p. 210.

$$= \frac{5k^3 - 27k^2\bar{x} + 27k\bar{x}^2}{6k^2 - 36k\bar{x} + 36\bar{x}^2}, \quad \frac{k}{3} \leq \bar{x} \leq \frac{2k}{3},$$

$$= \frac{3}{2}(k - \bar{x}), \qquad \frac{2k}{3} \leq \bar{x} \leq k.$$

Thus the regression curve of $W$ on $\bar{x}$ is continuous, but the regression is non-linear for $\frac{k}{3} \leq \bar{x} \leq \frac{2k}{3}$

3. The correlation between the arithmetic mean $\bar{x}$ and the median $\xi$.

Theorem II. *Let $f(x)$ be the probability function of the variable $x$. Let $F_2(\bar{x}, \xi)$ be that of the arithmetic mean $\bar{x}$ and the median $\xi$ in samples of three independent values of $x$. If $f(x)$ is a probability function of the first kind, then*

$$F_2(\bar{x}, \xi) = 18 f(\xi) \int_{3\bar{x}-2\xi}^{\infty} f(x_1) f(3\bar{x} - \xi - x_1) dx_1, \quad \xi \leq \bar{x},$$

$$= 18 f(\xi) \int_{\xi}^{\infty} f(x_1) f(3\bar{x} - \xi - x_1) dx_1, \quad \bar{x} \leq \xi.$$

Proof. Let $x_1$, $x_2$, $x_3$, be the three observed values of $x$. Write

$$x_1 + x_2 + x_3 = 3\bar{x},$$
$$x_2 = \xi$$
$$x_3 \leq x_2 \leq x_1$$

For $\bar{x}$ and $\xi$ assigned, $\xi \leq \bar{x}$, we must have

$$3\bar{x} - 2\xi \leq x_1 < \infty$$
$$x_2 = \xi$$
$$x_3 = 3\bar{x} - \xi - x_1,$$

and for $\bar{x} \leq \xi$,

$$\xi \leq x_1 < \infty$$
$$x_2 = \xi$$
$$x_3 = 3\bar{x} - \xi - x_1.$$

If we consider all possible arrangements of $x_1$ , $x_2$ , $x_3$ , we have

$$F_2(\bar{x},\xi)\,d\bar{x}\,d\xi = 6f(\xi)\,d\xi\int_{3\bar{x}-2\xi}^{\infty} f(x_1)\,f(x_3)\,dx_1\,dx_3, \qquad \xi \le \bar{x},$$

$$= 6f(\xi)\,d\xi\int_{\xi}^{\infty} f(x_1)\,f(x_3)\,dx_1\,dx_3, \qquad \bar{x} \le \xi.$$

The change of variable $x_3 = 3\bar{x} - \xi - x_1$ establishes the theorem.

In case of samples of five independent items $x_1$ , $x_2$ , $x_3$ , $x_4$ , $x_5$ , the probability function $F_2(\bar{x},\xi)$ is given by

$$F_2(\bar{x},\xi) = 150f(\xi)\int_{\xi}^{5\bar{x}-4\xi}\int_{5\bar{x}-3\xi-x_1}^{\infty}\int_{5\bar{x}-2\xi-x_1-x_2}^{\xi} f(x_1)f(x_2)f(x_4)f(5\bar{x}-\xi-x_1-x_2-x_4)\,dx_4\,dx_2\,dx_1$$

$$+150f(\xi)\int_{5\bar{x}-4\xi}^{\infty}\int_{\xi}^{\infty}\int_{5\bar{x}-2\xi-x_1-x_2}^{\xi} f(x_1)f(x_2)f(x_4)f(5\bar{x}-\xi-x_1-x_2-x_4)\,dx_4\,dx_2\,dx_1, \quad \xi \le \bar{x},$$

$$=150f(\xi)\int_{\xi}^{\infty}\int_{\xi}^{\infty}\int_{5\bar{x}-2\xi-x_1-x_2}^{\xi} f(x_1)f(x_2)f(x_4)f(5\bar{x}-\xi-x_1-x_2-x_4)\,dx_4\,dx_2\,dx_1, \quad \bar{x} \le \xi.$$

This follows immediately from the fact that for $\bar{x}$ and $\xi$ assigned, $\xi \le \bar{x}$ , we may have either

$$\begin{aligned} \xi \le x_1 &\le 5\bar{x} - 4\xi, \\ 5\bar{x} - 3\xi - x_1 \le x_2 &< \infty, \\ 5\bar{x} - 2\xi - x_1 - x_2 \le x_4 &\le \xi, \\ x_3 &= \xi, \\ x_5 &= 5\bar{x} - \xi - x_1 - x_2 - x_4, \end{aligned}$$

or

$$\begin{aligned} 5\bar{x} - 4\xi \le x_1 &< \infty, \\ \xi \le x_2 &< \infty, \\ 5\bar{x} - 2\xi - x_1 - x_2 \le x_4 &\le \xi \\ x_3 &= \xi \\ x_5 &= 5\bar{x} - \xi - x_1 - x_2 - x_4, \end{aligned}$$

and for $\bar{x} \leq \xi$ , we must have

$$\begin{aligned}
\xi \leq x_1 &< \infty \\
\xi \leq x_2 &< \infty, \\
5\bar{x} - 2\xi - x_1 - x_2 \leq x_4 &\leq \xi, \\
x_3 &= \xi \\
x_5 &= 5\bar{x} - \xi - x_1 - x_2 - x_4 .
\end{aligned}$$

If $f(x)$ is a probability function of the second kind, it is clear that $0 \leq \xi \leq \frac{3\bar{x}}{2}$ in samples of three items. Then

$$\begin{aligned}
F_2(\bar{x},\xi) &= 18 f(\xi) \int_{3\bar{x}-2\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\, dx_1, \qquad 0 \leq \xi \leq \bar{x}, \\
&= 18 f(\xi) \int_{\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\, dx_1, \qquad \bar{x} \leq \xi \leq \frac{3\bar{x}}{2}.
\end{aligned}$$

In case of samples of five independent items drawn at random from a universe characterized by a probability function of the second kind, $F_2(\bar{x},\xi)$ can best be expressed in a form employing the notation used previously. Thus we write

$$\Phi = f(x_1) f(x_2) f(x_4) f(5\bar{x}-\xi-x_1-x_2-x_4),$$

$$u_{ij} = 5\bar{x} - i\xi - x_1 - x_2 - \cdots - x_j,$$

and

$$\int_a^b \int_c^d \int_e^f \Phi \, dx_4\, dx_2\, dx_1 = \begin{pmatrix} b & d & f \\ a & c & e \end{pmatrix} \Phi .$$

Then

$$\begin{aligned}
F_2(\bar{x},\xi) = 150 f(\xi) \Bigg[ &\begin{pmatrix} u_{40} & u_{21} & \xi \\ \xi & u_{31} & u_{22} \end{pmatrix} \Phi + \begin{pmatrix} u_{40} & u_{11} & u_{12} \\ \xi & u_{21} & 0 \end{pmatrix} \cdot \Phi \\
&+ \begin{pmatrix} u_{30} & u_{21} & \xi \\ u_{40} & \xi & u_{22} \end{pmatrix} \Phi + \begin{pmatrix} u_{30} & u_{11} & u_{12} \\ u_{40} & u_{21} & 0 \end{pmatrix} \cdot \Phi \\
&+ \begin{pmatrix} u_{20} & u_{11} & u_{12} \\ u_{30} & \xi & 0 \end{pmatrix} \Phi \Bigg], \qquad 0 \leq \xi \leq \bar{x},
\end{aligned}$$

$$= 150 f(\xi)\left[\begin{pmatrix} u_{30} & u_{21} & \xi \\ \xi & \xi & u_{22} \end{pmatrix}\Phi + \begin{pmatrix} u_{30} & u_{11} & u_{12} \\ \xi & u_{21} & 0 \end{pmatrix}\Phi\right.$$

$$\left. + \begin{pmatrix} u_{20} & u_{11} & u_{12} \\ u_{30} & \xi & 0 \end{pmatrix}\Phi\right], \qquad \bar{x} \le \xi \le \frac{5\bar{x}}{4},$$

$$= 150 f(\xi)\left[\begin{pmatrix} u_{20} & u_{11} & u_{12} \\ \xi & \xi & 0 \end{pmatrix}\Phi\right], \qquad \frac{5\bar{x}}{4} \le \xi \le \frac{5\bar{x}}{3}.$$

Finally, consider $f(x)$ to be a probability function of the third kind. In samples of three independent items, for $0 \le \bar{x} \le k/3$, we obtain $0 \le \xi \le 3\bar{x}/2$ ; for $k/3 \le \bar{x} \le 2k/3$ , we obtain $(3\bar{x}-k)/2 \le \xi \le 3\bar{x}/2$; for $2k/3 \le \bar{x} \le k$ , we obtain $(3\bar{x}-k)/2 \le \xi \le k$. It is not difficult to verify for $\bar{x}$ and $\xi$ assigned as indicated, the following regions of selection of $x_1$ are valid:

for $0 \le \bar{x} \le k/3$ and $0 \le \xi \le \bar{x}$,
or for $k/3 \le \bar{x} \le k/2$ and $3\bar{x}-k \le \xi \le \bar{x}$ , then
$3\bar{x}-2\xi \le x_1 \le 3\bar{x}-\xi$;

for $k/3 \le \bar{x} \le k/2$ and $(3\bar{x}-k)/2 \le \xi \le 3\bar{x}-k$,
or for $k/2 \le \bar{x} \le k$ and $(3\bar{x}-k)/2 \le \xi \le \bar{x}$ , then
$3\bar{x}-2\xi \le x_1 \le k$;

for $0 \le \bar{x} \le k/2$ and $\bar{x} \le \xi \le 3\bar{x}/2$,
or for $k/2 \le \bar{x} \le 2k/3$ and $3\bar{x}-k \le \xi \le 3\bar{x}/2$ , then
$\xi \le x_1 \le 3\bar{x}-\xi$;

for $k/2 \le \bar{x} \le 2k/3$ and $\bar{x} \le \xi \le 3\bar{x}-k$,
or for $2k/3 \le \bar{x} \le k$ and $\bar{x} \le \xi \le k$ , then $\xi \le x_1 \le k$.

Thus

$$F_2(\bar{x},\xi) = 18 f(\xi)\int_{3\bar{x}-2\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\,dx_1,$$

$$= 18 f(\xi)\int_{3\bar{x}-2\xi}^{k} f(x_1) f(3\bar{x}-\xi-x_1)\,dx_1,$$

$$= 18 f(\xi) \int_{\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\, dx_1 ,$$

$$= 18 f(\xi) \int_{\xi}^{k} f(x_1) f(3\bar{x}-\xi-x_1)\, dx_1 ,$$

over those regions of the $\bar{x}\xi$-plane as indicated above.

With samples of five items, the correlation surface is defined in so many parts that we shall not take the space necessary to consider it.

As illustrations of these theorems, we shall find the correlation between the median and the mean for universes of specified types.

Example 1. Let $f(x) = e^{-x}, \quad 0 \leq x < \infty.$
For samples of three items, we have

$$F_2(\bar{x},\xi) = 18\xi e^{-3\bar{x}}, \quad 0 \leq \xi \leq \bar{x},$$
$$= 18(3\bar{x}-2\xi) e^{-3\bar{x}}, \quad \bar{x} \leq \xi \leq \frac{3\bar{x}}{2}$$

The distribution function of the marginal totals of $\xi$ is given by[4]

$$\varphi(\xi) = 6e^{-2\xi}(1-e^{-\xi}), \quad 0 \leq \xi < \infty .$$

For $\bar{x}$ assigned, the mean of the array of $\xi$ is

$$\bar{\xi}_{\bar{x}} = \frac{5\bar{x}}{6}$$

Thus the regression of $\xi$ on $\bar{x}$ is linear and $r = \frac{5\sqrt{267}}{89}.$

Example 2. Let $f(x) = \frac{1}{k}, \quad 0 \leq x \leq k.$
For samples of three items, we have

$$F_2(\bar{x},\xi) = \frac{18\xi}{k^3}$$

$$= \frac{18}{k^3}(k-3\bar{x}+2\xi),$$

[4]Cf. American Journal of Mathematics, Vol. 54 (1932), p. 364.

$$= \frac{18}{k^3}(3\bar{x}-2\xi),$$

$$= \frac{18}{k^3}(3k-\xi),$$

over those regions of the $\bar{x}\xi$-plane indicated above. The distribution function of the marginal totals of $\xi$ is given by[5]

$$\varphi(\xi) = \frac{6\xi}{k^3}(k-\xi), \qquad 0 \leq \xi \leq k.$$

We find

$$\begin{aligned} \bar{\xi}_{\bar{x}} &= \frac{5\bar{x}}{6}, & 0 \leq \bar{x} \leq \frac{k}{3}, \\ &= \frac{5\bar{x}^3-(3\bar{x}-k)^3}{6\bar{x}^2-2(3\bar{x}-k)^2}, & \frac{k}{3} \leq \bar{x} \leq \frac{2k}{3}, \\ &= \frac{(5\bar{x}+k)}{6}, & \frac{2k}{3} \leq \bar{x} \leq k. \end{aligned}$$

Thus the regression curve of $\xi$ on $\bar{x}$ is continuous but the regression is non-linear for $\frac{k}{3} \leq \bar{x} \leq \frac{2k}{3}$.

4. The correlation between the median $\xi$ and the range $W$.

Theorem III. *Let $f(x)$ be the probability function of the variable $x$. Let $F_3(\xi,W)$ be that of the median $\xi$ and the range $W$ in samples of $2m+1$ independent values of $x$. If $f(x)$ is a probability function of the first kind, then*

$$F_3(\xi,W) = \frac{(2m+1)!}{[(m-1)!]^2} f(\xi) \int_{\xi}^{\xi+W} f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1} f(t)\,dt\right]^{m-1}\left[\int_{x_1-W}^{\xi} f(t)\,dt\right]^{m-1} dx_1.$$

Proof. We have

$$x_1 - x_{2m+1} = W,$$

$$x_{m+1} = \xi,$$

$$\xi \leq x_2, \cdots, x_m \leq x_1,$$

$$x_1 - W \leq x_{m+1}, \cdots, x_{2m} \leq \xi.$$

[5]Cf. P. R. Rider, On the Distribution of the Ratio of Mean to Standard Deviation, etc., Biometrika, Vol. 21 (1929), pp. 136-137.

Hence the theorem.

If $f(x)$ is a probability function of the second kind, then

$$F_3(\xi,W)=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{\xi}^{\xi+W}f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1}f(t)dt\right]^{m-1}\left[\int_{x_1-W}^{\xi}f(t)dt\right]^{m-1}dx_1,\quad W\leq\xi,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{W}^{\xi+W}f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1}f(t)dt\right]^{m-1}\left[\int_{x_1-W}^{\xi}f(t)dt\right]^{m-1}dx_1,\quad \xi\leq W.$$

Finally, consider $f(x)$ to be a probability function of the third kind. We observe for $0\leq\xi\leq k$, that $0\leq W\leq k$. For assigned values of $\xi$ and $W$, the following regions of selection of $x_1$ are obvious:

for $0\leq\xi\leq k/2$, and $0\leq W\leq\xi$,
or for $k/2\leq\xi\leq k$ and $0\leq W\leq k-\xi$, then $\xi\leq x_1\leq\xi+W$;
for $0\leq\xi\leq k/2$ and $\xi\leq W\leq k-\xi$, then $W\leq x_1\leq\xi+W$;
for $0\leq\xi\leq k/2$, and $k-\xi\leq W\leq k$,
or for $k/2\leq\xi\leq k$ and $\xi\leq W\leq k$, then $W\leq x_1\leq k$;
for $k/2\leq\xi\leq k$ and $k-\xi\leq W\leq\xi$, then $\xi\leq x_1\leq k$.

If we write

$$\psi=f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1}f(t)dt\right]^{m-1}\left[\int_{x_1-W}^{\xi}f(t)dt\right]^{m-1},$$

we have

$$F_3(\xi,W)=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{\xi}^{\xi+W}\psi\,dx_1,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{W}^{\xi+W}\psi\,dx_1,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{W}^{k}\psi\,dx_1,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{\xi}^{k}\psi\,dx_1,$$

over those regions of the $\xi W$-plane previously indicated.

We shall consider two simple examples.

Example 1. Let $f(x) = e^{-x}$, $0 \leq x < \infty$.

With samples of three items,

$$F_3(\xi, W) = 3e^{-3\xi}(e^{W} - e^{-W}), \quad W \leq \xi,$$
$$= 3e^{-W-\xi}(1 - e^{-2\xi}), \quad \xi \leq W.$$

The regression is readily shown to be non-linear.

Example 2. Let $f(x) = \frac{1}{k}$, $0 \leq x \leq k$.

With samples of three items,

$$\begin{aligned} F_3(\xi, W) &= \frac{6W}{k^3}, \\ &= \frac{6\xi}{k^3}, \\ &= \frac{6}{k^3}(k - W) \\ &= \frac{6}{k^3}(k - \xi), \end{aligned}$$

over those regions of the $\xi W$-plane which have been previously given. The mean of the array of $W$ corresponding to an assigned $\xi$ is $\overline{W}_\xi = \frac{k}{2}$. Accordingly, there is no correlation between the median and the range in samples of three items drawn from this universe.

It is easy to employ the type of argument used in establishing Theorem III to obtain the probability function of the median and lower quartile. Thus, if $f(x)$ is a probability function of the second kind and $F_4(\xi, \eta)$ is the probability function of the median $\xi$ and the lower quartile $\eta$ in samples of $4m+1$ items, then

$$F_4(\xi, \eta) = \frac{(4m+1)!}{(2m)!\,m!\,(m-1)!} f(\xi) f(\eta) \left[\int_\xi^\infty f(t)\,dt\right]^{2m} \left[\int_0^\eta f(t)\,dt\right]^m \cdot \left[\int_\eta^\xi f(t)\,dt\right]^{m-1}, \quad \eta \leq \xi.$$

# ON THE DEGREE OF APPROXIMATION OF CERTAIN QUADRATURE FORMULAS

*By*

A. L. O'Toole
*National Research Fellow.*

If $f(x)$ be a continuous function of period $2\pi$, and if the interval under consideration, say the interval from $0$ to $2\pi$, be divided into $m$ equal parts by the $m+1$ points $x_i = 2i\pi/m$, $i = 0, 1, 2, \ldots, m$, then the trigonometric sum of the $n$th order coinciding in value with $f(x)$ at the $m+1$ points $x_i$, or the trigonometric sum of the $n$th order lacking the term in $\sin nx$, is, according as $m = 2n+1$ or $m = 2n$,

$$\Phi_n(x) = \tfrac{1}{2}a_0 + a_1 \cos x + a_2 \cos 2x + \cdots \cdots + a_n \cos nx$$
$$+ b_1 \sin x + b_2 \sin 2x + \cdots + b_n \sin nx$$

or

$$u_n(x) = \tfrac{1}{2}a_0 + a_1 \cos x + a_2 \cos 2x + \cdots \cdots + \tfrac{1}{2} a_n \cos nx$$
$$+ b_1 \sin x + b_2 \sin 2x + \cdots + b_{n-1} \sin(n-1)x,$$

where

$$a_k = \frac{h}{\pi} \sum_{i=1}^{m} f(x_i) \cos kx_i, \qquad h = \frac{2\pi}{m},$$

$$b_k = \frac{h}{\pi} \sum_{i=1}^{m} f(x_i) \sin kx_i.$$

If the Fourier coefficients of $f(x)$ be denoted by

$$\alpha_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx,$$

$$\beta_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx,$$

then it has been shown[1] that the interpolating coefficients $a_k$ and $b_k$ are approximations to the Fourier coefficients $\alpha_k$ and $\beta_k$ in the sense of the rectangle quadrature formula, in the sense of the trapezoid quadrature formula, in the sense of the average of the results of two applications of Simpson's formula, and in the sense of higher quadrature formulas. In other words, the simple rectangle formulas $a_k$ and $b_k$ are as good approximations to the areas $\alpha_k$ and $\beta_k$ as the estimates given by the trapezoid rule, the average of two applications of Simpson's rule, or higher quadrature formulas.

It is the purpose of this note to discuss certain quadrature formulas and to observe some other conditions under which the rectangle formula will give as good an approximation as the more complicated formulas.

The most elementary and best known of the formulas are the rectangle formula, the trapezoid formula, and Simpson's formula. Many of the more complex rules are the results of attempts by different investigators[2] to improve by various devices the approximations given by these three simple rules.

Suppose the area under consideration is bounded by the curve $y = f(x)$, the $x$-axis and the ordinates at $x = a$ and $x = b$. If the interval from $a$ to $b$ be devided into $n$ equal[3] parts, say of length $h$, by the $n+1$ points $x_0 = a, x_1, x_2, \cdots, x_{n-1}, x_n = b$, and if rectangles, each of width $h$ and height $y_i$, $i = 0, 1, 2, \cdots, n-1$, be constructed, then the area as approximated by these $n$ rectangles is

$$(1) \qquad A = h \sum_{\nu=0}^{n-1} y_\nu .$$

---

[1]D. Jackson, Some Notes on Trigonometric Interpolation, Amer. Math. Monthly, vol. xxxiii, no. 8, October 1927.

[2]See Runge and Willers, Encyklopädie Der Mathematischen Wissenschaften, Bd. II:3 (1915), pp. 45-176.

[3]Discussion from point of view of least squares, Otto Biermann, Monatshefte Fur Mathematik Und Physik, 14 (1903), pp. 226-242.

For unequal intervals see Jas. W. Glover, International Mathematical Congress, Toronto, 1924.

To find an expression for the error we assume the first derivative exists, so that for the first rectangle

$$f(x) = f(a) + (x-a)f'(u),$$

$$\int_a^{a+h} f(x)\,dx = hf(a) + \frac{h^2}{2}f'(z), \quad a < z < a+h.$$

Hence the error for the $n$ rectangles is

$$\text{(1e)} \qquad E = \frac{h^2}{2}\sum_{v=1}^{n} f'(z_v) = \frac{(b-a)^2}{2n} f'(z), \quad a < z < b,$$

i.e., an error of the order of $\frac{1}{n}$.

Let $n = mk$, $k = 1, 2, 3, \cdots$. If we approximate the area in the first $k$ subintervals by a parabola of degree $k$ coinciding in value with $f(x)$ at the first $k$ values of $x$, then integrating Lagrange's interpolation formula an expression for the error is obtained. If $k$ is odd then

$$E_1 = C_1\left(\frac{H}{2}\right)^{k+2} \frac{f^{(k+1)}(z)}{(n+1)!}, \quad H = kh,$$

where

$$C_1 = \frac{1}{(k+2)k^{k-1}} \int_{-1}^{1} (t^2-1)(k^2t^2-1^2)(k^2t^2-3^2)\cdots(k^2t^2-(k-2)^2)\,dt.$$

If $k$ is even, then making use of Rolle's Theorem,

$$E_2 = C_2\left(\frac{H}{2}\right)^{k+3} \frac{f^{(k+2)}(z)}{k+2!}$$

where

$$C_2 = \frac{1}{k^{k-2}} \int_{-1}^{1} (t^2-1)(t^2)(k^2t^2-2^2)\cdots(k^2t^2-(k-2)^2)\,dt.$$

The error over the whole interval will be obtained by summing the $m$ errors corresponding to each $k$ subintervals.

If $n$ trapezoids are formed by joining the ends of successive ordinates then the area as approximated by the sum of the areas of these trapezoids is

$$A = \frac{h}{2} \sum_{v=0}^{n-1} (y_v + y_{v+1}) \tag{2}$$

and the error is

$$E = -\frac{(b-a)^3}{12n^2} f''(z), \tag{2e}$$

i.e., an error of the order of $\frac{1}{n^2}$ .

Simpson's formula may be obtained by passing second degree parabolas through the ends of three successive ordinates, that is $k=2$ , and gives

$$A = \frac{h}{3}\left[2 \sum_{v=0}^{m} y_{2v} + 4\sum_{v=1}^{m} y_{2v-1} - (y_0 + y_{2m})\right], \quad n=2m. \tag{3}$$

The error is

$$E = -\frac{(b-a)^5}{180n^4} f^{iv}(z), \tag{3e}$$

i.e., an error of the order of $\frac{1}{n^4}$ .

To illustrate the fact that sometimes the rectangle formula (1) gives a better approximation than the Simpson formula (3) these formulas will be applied to the problem of finding the area under the so-called normal curve of error. From a table[4] giving five places of decimals it is seen that the ordinates to the right of $x = 4.76$ and to the left of $x = -4.76$ are everywhere zero if the equation be written in the form $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ . Divide the interval from $x = -4.80$ to $x = 4.80$ into eight partial intervals each of length 1.20. Formula (1) gives $A = .99998$ while

[4]Jas. W. Glover, Tables of Applied Mathematics in Finance, Insurance and Statistics.

(3) gives $A = .97834$, the same ordinates being used in each case.

There are three objections to the nature of Simpson's formula. They are the lack of smoothness at the points of intersection of the parabolas, the unequal weights attached to the odd and even numbered ordinates, and the requirement that the number of ordinates be odd.

Catalan[5] notices the lack of smoothness at the intersections of the parabolas used in setting up Simpson's rule and improves on it by passing parabolas through three successive ordinates and then retaining only the first half of each parabola except in the case of the last three ordinates where it is necessary to retain the whole parabola. To counterbalance the asymmetry introduced by these last three ordinates he repeats the process beginning with the last ordinat and then takes the arithmetic mean of the two results as his formula.
This gives

$$(4)\quad A = h\left[\sum_{v=0}^{n} y_v - \frac{5}{8}(y_0 + y_n) + \frac{1}{6}(y_1 + y_{n-1}) - \frac{1}{24}(y_2 + y_{n-2})\right].$$

And, of course, the error is still of the order of $\frac{1}{n^4}$. This formula has the additional advantage that it holds no matter whether $n$ is even or odd.

Similarly Crotti[6] showed that the different weights attached to the odd and even numbered ordinates in Simpson's formula ic a disadvantage. And Parmentier[7] by subtracting Simpson's formula from twice Catalan's obtained a formula in which the weights are the reverse of those in Simpson's. Mansion[8] gave an alternative derivation of Catalan's formula, his derivation requiring, however, an even number of ordinates.

[5]E. Catalan, Nouvelles Annales, 1re series (1851), pp. 412-415.

[6]Crotti, Il Politechnio 33 (1885), pp. 193-207.

[7]Parmentier, Association française pour l'avancement des sciences, Session Grenoble, 1882.

[8]Mansion, Supplement zu Mathesis 1 (1881).

Catalan's formula may be thought of as the rectangle formula plus three correctional terms involving the first three and the last three ordinates. In the case of an even number of ordinates a formula[9] involving only two such correctional terms and giving an approximation of the order of the error in the single trapezoid, i.e., of the order of $\frac{1}{n^3}$, the error in a single trapezoid of width $\frac{(b-a)}{n}$ being $\frac{-f''(z)(b-a)^3}{12n^3}$, can be obtained by applying Simpson's formula to the first $2m-1$ ordinates and approximating the remaining area by the trepezoid rule. Repeat the process from the opposite end and take the arithmetic mean of the two results as the quadrature formula. This gives

$$(5) \qquad a = h\left[\sum_{v=0}^{n} y_v - \frac{7}{12}(y_0 + y_n) + \frac{1}{12}(y_1 + y_{n-1})\right], \quad n = 2m-1.$$

It is the only formula with just two correctional terms which will give even this order of approximation in general because any change in the coefficients of these end ordinates will introduce in general an error of the order of the error in the rectangle formula for a single subinterval, i.e., an error of the order of $\frac{1}{n^2}$.

Another important quadrature formula is called the three-eighths rule and is obtained by passing third order parabolas through four successive ordinates. It may be written

$$(6) \qquad A = \frac{3h}{8}\left[2\sum_{v=0}^{m} y_{3v} + 3\sum_{v=0}^{m-1} y_{3v+1} + 3\sum_{v=0}^{m-1} y_{3v+2} - (y_0 + y_{3m})\right], \quad n = 3m.$$

The error is

$$(6e) \qquad E = -\frac{(b-a)^5}{400n^4} f^{iv}(z),$$

i.e., an error of the same order as the error corresponding to Simpson's formula. The error terms derived from the Lagrange

[9]Durand, Engineering News, Jan. 1894. J. Lipka, Graphical and Mechanical Computation, Part II, p. 226.

formula shows the advantage of using parabolas of even degree.

Besides the fact that the order of the error is the same as that in the case of Simpson's formula, this three-eighths formula has disadvantages similar to those mentioned in the case of Simpson's formula. There is still a lack of smoothness at the intersections of the parabolas; the weights attached to the ordinates are as undesirable as before; and the number of partial intervals must be a multiple of three.

It is possible however to do away with these disadvantages by proceeding as follows. Pass a third order parabola through the first four ordinates $y_0$, $y_1$, $y_2$, $y_3$. Retain only the area in the first two partial intervals. Pass a third order parabola through the four ordinates $y_1$, $y_2$, $y_3$, $y_4$ and retain only the area in the central interval. Proceed in this way retaining each time only the area in the central interval until the last four ordinates are reached where it will again be necessary to retain the area in two strips, viz., the last two partial intervals. The sum of these areas gives the required quadrature formula. It is

$$(7)\quad A = h\left[\sum_{v=0}^{n} y_v - \tfrac{3}{8}(y_0+y_n)+\tfrac{7}{24}(y_1+y_{n-1})-\tfrac{1}{6}(y_2+y_{n-2})+\tfrac{1}{24}(y_3+y_{n-3})\right].$$

This formula holds for any $n$ greater than or equal to three. From the point of view of the order of the error this formula is, as one would expect, no better than Catalan's formula. As a matter of fact formula (7) can be obtained from formula (4) by subtracting from (4) $\frac{h}{24}(\Delta^3 y_0' - \Delta^3 y_{n-3})$ a quantity which, in general, is of the order of $\frac{1}{n^4}$.

If $n=4m$ and fourth order parabolas are used in approximating the area in four successive partial intervals then the formula is

$$(8)\quad A=\frac{4h}{45}\left[7\sum_{v=0}^{m} y_{4v}+16\sum_{v=0}^{m-1} y_{4v+1}+6\sum_{v=0}^{m-1} y_{4v+2}+16\sum_{v=0}^{m-1} y_{4v+3}-\frac{7}{2}(y_0+y_{4m})\right].$$

The error is

(8e) $$E = -\frac{2(b-a)^7}{945 n^6} f^{vi}(z),$$

i.e., an error of the order of $\frac{1}{n^6}$.

Several modifications may be made to improve this formula. For instance if $n = 2m+1$ then apply the fourth degree parabola to the ordinates $y_0$, $y_1$, $y_2$, $y_3$, $y_4$ and retain only the area in the first three strips. Apply a fourth degree parabola to the ordinates $y_2$, $y_3$, $y_4$, $y_5$, $y_6$ and retain the area in the two central strips. And so on till in the final step it will be necessary to retain the area in the last three strips. Addition gives the formula

(9) $$A = \frac{h}{720}\Big[896\sum_{v=0}^{m} y_{2v} + 544\sum_{v=1}^{m} y_{2v-1} - 653(y_0+y_{2m}) + 374(y_1+y_{2m-1}) - 256(y_2+y_{2m-2}) + 106(y_3+y_{2m-3}) - 19(y_4+y_{2m-4})\Big]$$

A formula which holds for any $n$ may be obtained by passing a fourth degree parabola through $y_0$, $y_1$, $y_2$, $y_3$, $y_4$ and retaining only the area between $y_0$ and $y_2$. Pass a fourth degree parabola through $y_1$, $y_2$, $y_3$, $y_4$, $y_5$ and retain only the area between $y_2$ and $y_3$. And so on, retaining only the area in one strip, until at the end it will be necessary to retain the area in the last three strips. Repeat the process beginning at the last ordinate and take the arithmetic mean. The result is

(10) $$A = h\Big[\sum_{v=0}^{n} y_v - \frac{193}{288}(y_0+y_n) + \frac{77}{240}(y_1+y_{n-1}) - \frac{7}{30}(y_2+y_{n-2}) + \frac{73}{720}(y_3+y_{n-3}) - \frac{3}{160}(y_4+y_{n-4})\Big].$$

This formula can be obtained in the case of an even number of ordinates by retaining three strips at the beginning, two from

then on, reversing the process and taking the arithmetic mean.

Formulas (4), (5), (7) and (10) not only give, in general, at least as good approximations as Simpson's formula, the trapezoid formula, the three-eighths formula, and the fourth degree formula (8) respectively, but in addition have the important property that under certain conditions they show that the simple rectangle formula must give at least as good an approximation as the higher formulas. If $f(x)$ is a function such that the curve $y = f(x)$ actually, or at least for practical purposes, coincides with the $x$-axis to the left of $x=a$ and to the right of $x=b$, then in dividing the interval from $a$ to $b$ into $h$ equal parts each of length $h$ it will not affect the area required if two, one, three or four partial intervals of length $h$ are marked off to the left of $a$ and to the right of $b$, the number of such partial intervals corresponding to (4), (5), (7) and (10) respectively. Hence it is seen that under these conditions (4), (5), (7) and (10) reduce to the simple rectangle formula (1).

If the curve coincides with the $x$-axis at one end of the interval over which the area is required but does not at the other end then formulas (4), (5), (7) and (10) become respectively

$$\text{(4a)}\quad A = h\Big(\sum_{v=0}^{n} y_v - \frac{5}{8}y_n + \frac{1}{6}y_{n-1} - \frac{1}{24}y_{n-2}\Big),$$

$$\text{(5a)}\quad A = h\Big(\sum_{v=0}^{2m-1} y_v - \frac{7}{12}y_{2m-1} + \frac{1}{12}y_{2m-2}\Big),$$

$$\text{(7a)}\quad A = h\Big(\sum_{v=0}^{n} y_v - \frac{2}{3}y_n + \frac{7}{24}y_{n-1} - \frac{1}{6}y_{n-2} + \frac{1}{24}y_{n-3}\Big),$$

$$\text{(10a)}\quad A = h\Big(\sum_{v=0}^{n} y_v - \frac{193}{288}y_n + \frac{77}{240}y_{n-1} - \frac{7}{30}y_{n-2} + \frac{73}{720}y_{n-3} - \frac{3}{160}y_{n-4}\Big).$$

For example, consider again the normal curve of error and suppose that the area to the left of the ordinate at $x=o$ is required. Formulas (4a), (5a), (7a) and (10a) apply and for sixteen par-

tial intervals give respectively $A=.49994$, $A=.49550$, $A=.50008$, and $A=.50002$, an extra partial interval to the left of $x=-480$ being used in the case of (5a) in order to have an odd number of intervals for that formula. Using thirty-two partial intervals the same formulas give $A=.49999$, $A=.49949$, $A=.50000$, and $A=.50000$ respectively.

If, as often happens, the values of ordinates outside the interval over which the area is required are known then even better quadrature formulas may be obtained. For example, suppose that in deriving formula (7) the ordinate $y_{-1}$ at a distance of $h$ to the left of $y_0$ and the ordinate $y_{n+1}$ at a distance $h$ to the right of $y_n$ are known. Then it will not be necessary to retain the areas in double strips at the beginning and end of the interval, and the formula for the area over the interval from $x=a$ to $x=b$ is

$$(11)\quad A=h\left[\sum_{\nu=0}^{n} y_\nu -\frac{1}{24}(y_{-1}+y_{n+1})-\frac{1}{2}(y_0+y_n)+\frac{1}{24}(y_1+y_{n-1})\right].$$

It should be noted that in case $y_{-1}$ and $y_{n+1}$ are known Catalan's formula reduces to (11). And, similarly, in the case of the derivation of formula (10) it will be necessary to retain the area in a single strip each time except in the case of the last application of the fourth degree parabola when it will be necessary to retain the area in the two central strips. The formula arrived at is

$$(12)\quad A=h\left[\sum_{\nu=0}^{n} y_\nu -\frac{3}{160}(y_{-1}+y_{n+1})-\frac{83}{144}(y_0+y_n)+\frac{2}{15}(y_1+y_{n-1})\right.$$
$$\left.-\frac{11}{240}(y_2+y_{n-2})+\frac{11}{1440}(y_3+y_{n-3})\right].$$

Formulas (11) and (12) reduce to the rectangle formula (1) under the same conditions as in the cases of (4), (5), (7) and (10). Likewise when the curve coincides with the $x$-axis to the left of $x=a$ (11) and (12) become

(11a) $A = h\left(\sum_{v=0}^{n} y_v - \frac{1}{24} y_{n+1} - \frac{1}{2} y_n + \frac{1}{24} y_{n-1}\right)$ , and

(12a) $A = h\left(\sum_{v=0}^{n} y_v - \frac{3}{160} y_{n+1} - \frac{83}{144} y_n + \frac{2}{15} y_{n-1} - \frac{11}{240} y_{n-2} + \frac{11}{1440} y_{n-3}\right)$

If we apply formula (11a) to finding the area under the normal curve to the left of the ordinate at $x=o$ and take $n=4$, $h=1.20$, $a=-4.80$ then we find $A=.49999$. In other words, in this case (11a) gives as good a result with six ordinates as (4a) or (7a) give with thirty-three ordinates or (6a) with thirty-four ordinates.

Quadrature formulas involving parabolas of degree higher than four have been obtained but they are to be used with caution on account of the great freedom they allow the approximating curves. However, modifications similar to those in this paper could also be made for these higher formulas. And the effect of any number of ordinates outside the ends of the interval could be noted.

This note will be concluded with a remark on the effect of errors in the data giving the values of the ordinates. Suppose the quadrature formula is $A = h(a_0 y_0 + a_1 y_1 + a_2 y_2 + \cdots\cdots + a_n y_n)$ and suppose further that each $y_i$ is subject to an error $e_i$, $i = 0, 1, 2, 3, \cdots, n$. If $e$ is the greatest of the absolute values of the $e_i$ then the error in $A$ cannot be greater than $he(a_0 + a_1 + a_2 + \cdots + a_n)$ if $a_0, a_1, a_2, \cdots\cdots, a_n$ are all positive, as will be true if parabolas of the fourth degree or lower are used. But $h(a_0 + a_1 + a_2 + \cdots + a_n) = (b-a)$ if the area is to be four from $x=a$ to $x=b$. Hence the error in $A$ due to errors in the data is not greater than $e(b-a)$. When parabolas of degree higher than four are used the coefficients in the quadrature formula are not always positive.

Vol. IV

# THE ANNALS
of
# MATHEMATICAL STATISTICS

*SIX DOLLARS PER ANNUM*

AUGUST 1933

PUBLISHED QUARTERLY BY

AMERICAN STATISTICAL ASSOCIATION

*Publication Office*—Edwards Brothers, Inc., Ann Arbor, Michigan

*Business Office*—530 Commerce Bldg., New York Univ., New York, N. Y.

*Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.*

STATEMENT OF OWNERSHIP

UNDER ACT OF CONGRESS OF AUGUST 24, 1912

*Publisher*—American Statistical Association, New York City, New York.

*Editor*—H. C. Carver, University of Michigan.

*Managing Editor*—Rattan S. Sekhon, Ann Arbor, Michigan.

*Business Manager*—J. W. Edwards, Ann Arbor, Michigan.

*Owner*—American Statistical Association, 530 Commerce Bldg., New York City.

# POLYNOMIAL APPROXIMATION BY THE METHOD OF LEAST SQUARES

*By* H. T. Davis

1. *Introduction.* In an earlier article in the Annals of Mathematics the author in collaboration with V. V. Latshaw published formulas and tables for the fitting of polynomials to data by the method of least squares.[1] In that paper two ranges of the independent variable were considered, one from $x=1$ to $x=p$, and the other from $x=-p$ to $x=p$. For the first range formulas were given for fitting polynomials of first, second and third degrees to data and these formulas were reduced to tables. For the second range formulas were given for polynomials from the first to the seventh degrees, but these formulas were not then reduced to tables.

It is the purpose of the present paper to supply the tables for the second range and hence to furnish a means of reducing to a minimum the numerical labor involved in fitting to data polynomials from the first to the seventh degree inclusive. Incidentally some novel mathematical aspects of the problem of polynomial approximation have been brought to light, particularly as it applies to the existence of a set of polynomials which are orthogonal for a summation over discrete intervals.

The tables have been computed in the statistical laboratory of Indiana University and have been checked by duplicate calculation. The computation has been made possible by grants of funds by the Waterman Institute of the University. The author is particularly indebted to Dr. V. V. Latshaw, Miss Irene Price, Byron Shelley, George Davis, and Miss Anna Lescisin for the work which they have done in connection with the various computations of this paper.

---

[1] Volume 31 (1930), pp. 52-78.

2. *Formulas.* Let us first consider the data to be given as a set of equally spaced items:

| $y$ | $y_1$ | $y_2$ | $y_3$ | $\cdot\ \cdot\ \cdot$ | $y_m$ |
|---|---|---|---|---|---|
| $x$ | $x_1$ | $x_2$ | $x_3$ | $\cdots\ \cdot\ \cdot$ | $x_m$ |

in which we assume that the difference $x_{i+1} - x_i$ is constant.

If $m$ is an odd number, $m = 2p+1$, we select zero as the center of the $x$-range and without loss of generality we replace the table just given by the following:

| $y$ | $y_{-p}$ | $y_{-p+1},$ | $\cdot\ \cdot$ | $y_{-1}$ | $y_0$ | $y_1$ | $\cdot\ \cdot$ | $y_{p-1}$ | $y_p$ |
|---|---|---|---|---|---|---|---|---|---|
| $x$ | $-p$ | $-p+1,$ | $\cdots$ | $-1$ | $0$ | $1$ | $\cdots$ | $p-1$ | $p$ |

Let us designate by $M_r$ the moments,

$$M_r = \sum_{s=-p}^{p} s^r y_s\,, \quad r = 0, 1, 2, \cdots, m. \tag{1}$$

If $m$ is an even number, $m = 2p$, we must make a slight change in the notation and consider the distribution,

| $y$ | $y_{-p}$ | $\cdot\ \cdot$ | $y_{-2}$ | $y_{-1}$ | $y_1$ | $y_2$ | $\cdot\ \cdot$ | $y_p$ |
|---|---|---|---|---|---|---|---|---|
| $x$ | $-(2p-1)/2$ | $\cdots$ | $-3/2$ | $-1/2$ | $1/2$ | $3/2$ | $\cdots$ | $(2p-1)/2$ |

The $r$-$th$ moments, $M_r'$, will be correspondingly equal to,

$$M_r' = (\tfrac{1}{2})^r \sum_{s=1}^{p} (2s-1)^r \left\{ y_s + (-1)^r y_{-s} \right\}.$$

The method of least squares is then employed as described in the previous paper to determine the coefficients of the polynomial,

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m, \quad m = 1, 2, \cdots, 7. \tag{2}$$

It will be unnecessary to repeat the explicit formulas obtained since they have been given in the previous paper, but it will be useful in explanation of the notation of the tables to give the fol-

lowing determination of the coefficients as linear functions of the moments:[2]

1. *The straight line,* $y = a_0 + a_1 x$.

(3) $a_0 = \dfrac{M_0}{(2p+1)} = AM_0$ , $a_1 = \dfrac{3M_1}{p(p+1)(2p+1)} = A'M_1$.

2. *The parabola,* $y = a_0 + a_1 x + a_2 x^2$

(4) $a_0 = AM_0 + BM_2$ , $a_2 = BM_0 + CM_2$ , $a_1$ determined from (3).

3. *The cubic,* $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$.

(5) $a_1 = A'M_1 + B'M_3$ , $a_3 = B'M_1 + C'M_3$, $a_0$ and $a_2$ determined from (4).

4. *The quartic,* $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$.

(6) $a_0 = AM_0 + BM_2 + CM_4$, $a_2 = BM_0 + DM_2 + EM_4$, $a_4 = CM_0 + EM_2 + FM_4$, $a_1$ and $a_3$ determined from (5).

5. *The quintic,* $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$

(7) $a_1 = A'M_1 + B'M_3 + C'M_5$, $a_3 = B'M_1 + D'M_3 + E'M_5$, $a_5 = C'M_1 + E'M_3 + F'M_5$, $a_0, a_2$ and $a_4$ determined from (6).

6. *The sextic,* $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6$

[2] The notation follows that of the previous paper. It should be noted that the coefficients $A$, $B$, $C$, etc. for the straight line, the parabola, the quartic, and the sextic, and the coefficients $A'$, $B'$, $C'$, etc. for the cubic, the quintic, and the septic are all given by different formulas, but it is hoped that the omission of subscripts denoting the degrees of the polynomials will lead to no confusion.

$$
(8)\quad
\begin{aligned}
a_0 &= AM_0 + BM_2 + CM_4 + DM_6,\\
a_2 &= BM_0 + EM_2 + FM_4 + GM_6,\\
a_4 &= CM_0 + FM_2 + HM_4 + IM_6,\\
a_6 &= DM_0 + GM_2 + IM_4 + JM_6,
\end{aligned}
$$

$a_1$, $a_3$ and $a_5$ determined from (7).

7. *The septic*, $y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$ $a_5x^5 + a_6x^6 + a_7x^7$.

$$
(9)\quad
\begin{aligned}
a_1 &= A'M_1 + B'M_3 + C'M_5 + D'M_7,\\
a_3 &= B'M_1 + E'M_3 + F'M_5 + G'M_7,\\
a_5 &= C'M_1 + F'M_3 + H'M_5 + I'M_7,\\
a_7 &= D'M_1 + G'M_3 + I'M_5 + J'M_7,
\end{aligned}
$$

$a_0$, $a_2$, $a_4$ and $a_6$ determined from (8).

3. *Orthogonal Polynomials*. In a paper the significance of which has perhaps never been fully appreciated, J. P. Gram investigated the problem of polynomial approximation over discrete intervals by means of orthogonal polynomials.[3] This method has since been more fully investigated by Edward Condon[4] and his work was made the basis of a method for obtaining least squares polynomials by R. T. Birge and J. D. Shea.[5] The work of the latter, however, while effecting a simplification, does not reduce the problem to its simplest form.

In a recent paper issued by the Hungarian National Committee on Economic Statistics, Karl Jordan has employed orthogonal functions in connection with binomial moments and has very

[3]Über die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate. Journal für Math., vol. 94 (1883), pp. 41-73.

[4]The Rapid Fitting of a Certain Class of Empirical Formulae by the Method of Least Squares. Univ. of California Publications in Mathematics, vol. 2 (1927), pp. 55-66.

[5]A Rapid Method for Calculating the Least Square Solution of a Polynomial of any Degree. *Ibid.*, pp. 67-118.

greatly simplified the numerical work of curve fitting.[6] The polynomials which he employs, however, appear in the form,

$$y = a_0 + a_1 x + a_2 x(x-1) + a_3 x(x-1)(x-2) + \cdots,$$

although in the final result they are numerically equivalent to the polynomials of the present paper.

Let us begin with a set of polynomials,

$$\phi_0(x), \phi_1(x), \phi_2(x), \cdots, \phi_m(x),$$

of degrees $0, 1, 2, \cdots, m$ respectively such that,

$$\sum_{x=-p}^{p} \phi_s(x)\, \phi_t(x) = 0, \quad \text{for} \quad s \neq t.$$

Assuming the existence of such a set of polynomials we can approximate by means of them a function $f(x)$ which is defined over the set of integers from $-p$ to $p$.

Writing the approximation equation,

$$(10) \quad f(x) = A_0\, \phi_0(x) + A_1\, \phi_1(x) + \cdots + A_m\, \phi_m(x),$$

we multiply by $\phi_r(x)$ and sum from $-p$ to $p$. We then obtain,

$$(11) \quad \sum_{-p}^{p} f(x)\, \phi_r(x) = A_r\, S_r \quad , \text{where } S_r = \sum_{-p}^{p} \phi_r^2(x).$$

If we represent the polynomial $\phi_r(x)$ by the series,

$$\phi_r(x) = \phi_0 + \phi_1 x + \phi_2 x^2 + \cdots + \phi_r x^r,$$

[6]See Berechnung der Trendlinie auf Grund der Theorie der kleinsten Quadrate, Budapest (1930) and Praktische Anwendung der Trendberechnungs-Methode von Jordan, by A. Sipos, Budapest (1930).

it is clear from the definition of the moments (1) that we get from (10) the evaluation,

$$A_r = \sum_{s=1}^{r} \phi_s M_s / S_r . \tag{12}$$

That these coefficients are identifiable with those explicitly given in equations (3) to (9) is a consequence of the following consideration:

Let us approximate $f(x)$ by minimizing the following sum:

$$J = \sum_{x=-p}^{p} \left[ f(x) - A_0 \phi_0(x) - A_1 \phi_1(x) - \cdots - A_m \phi_m(x) \right]^2 ,$$

which is equivalent in its result to the somewhat different method employed in the actual determination of the formulas (3) to (9).

Taking the derivative of $J$ with respect to $A_r$ and equating the result to zero we get,

$$2 \sum_{-p}^{p} \left[ f(x) \phi_r(x) - \sum_{s=0}^{m} A_s \phi_r(x) \phi_s(x) \right] = 0 ,$$

whence, recalling the orthogonality of the polynomials,

$$A_r = \sum_{-p}^{p} f(x) \phi_r(x) / S_r . \tag{13}$$

We thus see that the ratios $\phi_s / S_r$ can be written down explicitly by comparing them with the corresponding coefficients of $M_s$ in equations (3) to (9).

In particular, if $r = m$, we find the coefficients of the polynomial $\phi_m(x)$ by equating the right member of (13) with the corresponding last row in the formulas (3) to (9). For example,

if $r=5$, we have,

$$\frac{(\phi_1 M_1+\phi_2 M_2+\phi_3 M_3+\phi_4 M_4+\phi_5 M_5)}{S_5}=C'M_1+E'M_3+F'M_5$$

Hence we get,

$$\phi_1=C'S, \quad \phi_3=E'S, \quad \phi_5=F'S, \quad \phi_2=\phi_4=0.$$

By means of this identification we obtain as the first seven polynomials the following:[7]

$$\phi_0(x)=1/(2p+1)=A,$$

$$\phi_1(x)=3x/p(p+1)(2p+1)=A'x,$$

$$\phi_2(x)=\left\{\frac{3^2\,5^2}{p(p+1)(4p^2-1)(2p+3)}\right\}\left\{x^2-\frac{p(p+1)}{3}\right\}=Cx^2+B,$$

$$\phi_3(x)=\left\{\frac{5^2\,7}{p(p^2-1)(4p^2-1)(2p+3)(p+2)}\right\}\left\{x^3-\frac{(3p^2+3p-1)x}{5}\right\}=C'x^3+B'x,$$

$$\phi_4(x)=\left\{\frac{15^2\cdot 7^2}{4p(p^2-1)(4p^2-1)(4p^2-9)(2p+5)(p+2)}\right\}\cdot$$

$$\left\{x^4-\frac{(6p^2+6p-5)x^2}{7}+\frac{3p(p^2-1)(p+2)}{35}\right\}=Fx^4+Ex^2+C,$$

$$\phi_5(x)=\left\{\frac{3^4\cdot 7^2\,11}{4p(p^2-1)(4p^2-1)(4p^2-9)(p^2-4)(2p+5)(p+3)}\right\}\cdot$$

$$\left\{x^5-\frac{5(2p^2+2p-3)x^3}{9}+\frac{(15p^4+30p^3-35p^2-50p+12)x}{63}\right\}$$

$$=F'x^5+E'x^3+C'x,$$

$$\phi_6(x)=\left\{\frac{3^2\,7^2\,11^2\,13}{4p(p^2-1)(p^2-4)(4p^2-9)(4p^2-25)(p+3)(2p+7}\right\}\cdot$$

$$\left\{x^6-\frac{5(3p^2+3p-7)x^4}{11}+\frac{(5p^4+10p^3-20p^2-25p+14)x^2}{11}\right.$$

$$\left.-\frac{5p(p^2-1)(p^2-4)(p+3)}{3\;7\;11}\right\}=Jx^6+Ix^4+Gx^2+D,$$

$$\phi_7(x)=\left\{\frac{3^3\,5\,11^2\,13^2}{4p(p^2-1)(p^2-4)(p^2-9)(4p^2-1)(4p^2-9)(4p^2-25)(p+4)(2p+7)}\right\}\cdot$$

$$\left\{x^7-\frac{7(3p^2+3p-10)x^5}{13}+\frac{7(15p^4+30p^3-90p^2-105p+101)x^3}{11\cdot 13}\right.$$

$$\left.-\frac{35p^6+105p^6-280p^4-735p^3+497p^2+882p-180)x}{3\;11\cdot 13}\right\},$$

$$=J'x^7+I'x^5+G'x^3+D'x$$

In order to effect the computation of the sum $S_m=\sum_{-p}^{p}\phi_m^2(x)$

---

[7] See note at end of this section

we replace the value of $A_m$ as given by (13) in (10) and compare the coefficient of $x^m$ with the corresponding coefficient in the proper formula of the set from (3) to (9). Thus for $m=5$, since $\phi_5(x) = F'x^5 + E'x^3 + C'x,$ we have from the quintic approximation,

$$f(x) = x^5(F'M_5 + E'M_3 + C'M_1) \quad + \text{ terms of lower degree,}$$

$$= F'x^5\left(\frac{\phi_5 M_5}{S_5} + \frac{\phi_3 M_3}{S_5} + \frac{\phi_1 M_1}{S_5}\right) \quad + \text{ terms of lower degree,}$$

$$= \left(\frac{F'^2 M_5}{S_5} + \frac{F'E'M_3}{S_5} + \frac{F'C'M_1}{S_5}\right)x^5 + \text{ terms of lower degree,}$$

Equating the coefficients of $M_5$ it is clear that $S_5 = F'$.

4. *The Recursion Formula.* It will be obvious from the preceding discussion that the polynomials which we have investigated are essentially the analogue of the well-known Legendre polynomials, where the integration between the limits $-1$ and $+1$ used in the definition of the latter's orthogonality is here replaced by the discrete sum over the integers from $-p$ to $+p$. We might, therefore, expect to find a recursion formula connecting any successive three of the new polynomials similar to the recursion formula which exists for the Legendre case. It turns out that this expectation is justified and we find the following relationship holding between $\phi_{n+1}(x)$, $\phi_n(x)$, and $\phi_{n-1}(x)$:

$$\begin{aligned}(14)\quad & (n+1)^2(2p-n)(2p+n+2)\phi_{n+1}(x) \\ & -4(2n+1)(2n+3)x\,\phi_n(x) \\ & +4(2n+1)(2n+3)\,\phi_{n-1}(x) = 0\end{aligned}$$

From this equation we easily deduce that the coefficient, $\phi_{n+1}$, of $x^{n+1}$ in $\phi_{n+1}(x)$ is related to the coefficient, $\phi_n$, of $x^n$ in

$\phi_n(x)$ as follows:

$$\phi_{n+1} = 4(2n+1)(2n+3)\phi_n / (n+1)^2(2p-n)(2p+n+2).$$

From this we obtain by iteration and a proper change in notation, the following value for $\phi_n$:

$$\phi_n = \{4^n 1^2 3^2 5^2 \ldots (2n-1)^2(2n+1)\} / \{(n!)^2 2p(4p^2-1)\cdot (4p^2-4)(4p^2-9)\ldots[4p^2-(n-1)^2](2p+n)(2p+n+1)\}. \tag{15}$$

The value thus obtained is at once seen to be equal to

$$S_n = \sum_{-p}^{p} \phi_n^2(x).$$

As an example consider the case where $n=2$. We then have,

$$\phi_2^2(x) = C^2x^4+2CBx^2+B^2 = a^2\{x^4 - \tfrac{2}{3}p(p+1)x^2 \quad + p^2(p+1)^2/9\},$$

where we abbreviate, $a^2 = 3^4 \cdot 5^2/p^2(p+1)^2(4p^2-1)^2(2p+3)^2$

We then obtain,

$$\sum_{-p}^{p}\phi_2^2(x) = a^2\{\sum_{-p}^{p}x^4 - 2p(p+1)\sum_{-p}^{p}\tfrac{x^2}{3} + p^2(p+1)^2(2p+1)/9\},$$

$$= a^2\{\tfrac{p}{15}(p+1)(2p+1)(3p^2+3p-1) \quad -\tfrac{2p^2(p+1)^2(2p+1)}{9} + p^2(p+1)^2(2p+1)/9\},$$

$$= a^2\{p(p+1)(4p^2-1)(2p+3)/45\} = 3^2 5/p(p+1)(4p^2-1)(2p+3)$$

which is seen to agree with the value of $S_2$ as calculated directly from (15).[8]

[8]The definition of the $\phi_n(x)$ which we have given above was chosen for the obvious connection which the functions in that form have with the problem of curve fitting and with the computed values in the tables. If, however, the coefficient of $x^n$ were reduced to unity,

$$\phi_0(x)=1,\ \phi_1(x)=x,\ \phi_2(x)=x^2-p(p+1)/3,$$

etc., then the recursion formula (14) would have been,

$$4(4n^2-1)\phi_{n+1}(x) - 4(4n^2-1)x\phi_n(x) + n^2(2p-n+1)(2p+n+1)\phi_{n-1}(x) = 0$$

If, moreover, $\phi_n(x)$ as just defined were multiplied by the coefficient $1\ 3\ 5 \ldots (2n-1)/n!$, which is the multiplier of the corresponding Legendre polynomials, then the recursion formula becomes,

$$4(n+1)\phi_{n+1}(x) - 4(2n+1)\phi_n(x) + n(2p-n+1)(2p+n+1)\phi_{n-1}(x) = 0$$

we replace the value of $A_m$ as given by (13) in (10) and compare the coefficient of $x^m$ with the corresponding coefficient in the proper formula of the set from (3) to (9). Thus for $m=5$, since $\phi_5(x) = F'x^5 + E'x^3 + C'x,$ we have from the quintic approximation,

$$f(x) = x^5(F'M_5 + E'M_3 + C'M_1) \quad + \text{ terms of lower degree,}$$

$$= F'x^5\left(\frac{\phi_5 M_5}{S_5} + \frac{\phi_3 M_3}{S_5} + \frac{\phi_1 M_1}{S_5}\right) \quad + \text{ terms of lower degree,}$$

$$= \left(\frac{F'^2 M_5}{S_5} + \frac{F'E'M_3}{S_5} + \frac{F'C'M_1}{S_5}\right)x^5 + \text{ terms of lower degree,}$$

Equating the coefficients of $M_5$ it is clear that $S_5 = F'$.

4. *The Recursion Formula.* It will be obvious from the preceding discussion that the polynomials which we have investigated are essentially the analogue of the well-known Legendre polynomials, where the integration between the limits $-1$ and $+1$ used in the definition of the latter's orthogonality is here replaced by the discrete sum over the integers from $-p$ to $+p$. We might, therefore, expect to find a recursion formula connecting any successive three of the new polynomials similar to the recursion formula which exists for the Legendre case. It turns out that this expectation is justified and we find the following relationship holding between $\phi_{n+1}(x)$, $\phi_n(x)$, and $\phi_{n-1}(x)$:

$$\begin{aligned} &(n+1)^2(2p-n)(2p+n+2)\phi_{n+1}(x) \\ (14) \quad &-4(2n+1)(2n+3)x\,\phi_n(x) \\ &+4(2n+1)(2n+3)\,\phi_{n-1}(x) = 0. \end{aligned}$$

From this equation we easily deduce that the coefficient, $\phi_{n+1}$, of $x^{n+1}$ in $\phi_{n+1}(x)$ is related to the coefficient, $\phi_n$, of $x^n$ in

$\phi_n(x)$ as follows:

$$\phi_{n+1} = 4(2n+1)(2n+3)\phi_n / (n+1)^2(2p-n)(2p+n+2).$$

From this we obtain by iteration and a proper change in notation, the following value for $\phi_n$:

$$(15)\quad \phi_n = \{4^n 1^2 3^2 5^2 \ldots (2n-1)^2(2n+1)\} / \{(n!)^2 2p(4p^2-1) \cdot (4p^2-4)(4p^2-9) \ldots [4p^2-(n-1)^2](2p+n)(2p+n+1)\}.$$

The value thus obtained is at once seen to be equal to

$$S_n = \sum_{-p}^{p} \phi_n^2(x).$$

As an example consider the case where $n=2$. We then have,

$$\phi_2^2(x) = C^2x^4 + 2CBx^2 + B^2 = a^2\{x^4 - \tfrac{2}{3}p(p+1)x^2 + p^2(p+1)^2/9\},$$

where we abbreviate, $a^2 = 3^4 \cdot 5^2 / p^2(p+1)^2(4p^2-1)^2(2p+3)^2$.

We then obtain,

$$\sum_{-p}^{p} \phi_2^2(x) = a^2\{\sum_{-p}^{p} x^4 - 2p(p+1)\sum_{-p}^{p} \frac{x^2}{3} + p^2(p+1)^2(2p+1)/9\},$$

$$= a^2\{\frac{p}{15}(p+1)(2p+1)(3p^2+3p-1) - \frac{2p^2(p+1)^2(2p+1)}{9} + p^2(p+1)^2(2p+1)/9\},$$

$$= a^2\{p(p+1)(4p^2-1)(2p+3)/45\} = 3^2 5 / p(p+1)(4p^2-1)(2p+3)$$

which is seen to agree with the value of $S_2$ as calculated directly from (15).[8]

[8]The definition of the $\phi_n(x)$ which we have given above was chosen for the obvious connection which the functions in that form have with the problem of curve fitting and with the computed values in the tables. If, however, the coefficient of $x^n$ were reduced to unity,

$$\phi_0(x) = 1,\ \phi_1(x) = x,\ \phi_2(x) = x^2 - p(p+1)/3,$$

etc., then the recursion formula (14) would have been,

$$4(4n^2-1)\phi_{n+1}(x) - 4(4n^2-1)x\phi_n(x) + n^2(2p-n+1)(2p+n+1)\phi_{n-1}(x) = 0$$

If, moreover, $\phi_n(x)$ as just defined were multiplied by the coefficient $1 \cdot 3\,5 \ldots (2n-1)/n!$, which is the multiplier of the corresponding Legendre polynomials, then the recursion formula becomes,

$$4(n+1)\phi_{n+1}(x) - 4(2n+1)\phi_n(x) + n(2p-n+1)(2p+n+1)\phi_{n-1}(x) = 0$$

5. *The Polynomials of Gram.* It will be at once evident that the results obtained above permit us to define a new set of polynomials which are orthogonal over the discrete range from

$$x=1 \text{ to } x=p'$$

In the former paper in the Annals it was proved that the formulas for the coefficients of the least square polynomial,

$$y=A_0+A_1x+\cdots\cdots+A_m x^m,$$

fitted to data given over the discrete range $x=1$ to $x=p'$, can be obtained from the coefficients, $a_0, a_1, a_2, \cdots, a_m$, of equation (2), by means of the following substitution:

$$p=(p'-1)/2,$$

$$M_r=m_r-r\left(\frac{p'+1}{2}\right)m_{r-1}+\frac{r(r-1)}{2!}\left(\frac{p'+1}{2}\right)^2 m_{r-2}-\cdots\cdots+(-1)^r\left(\frac{p'+1}{2}\right)^r m_0,$$

where $M_r$ are the moments defined by (1) and $m_r$ are the moments,

$$m_r=\sum_{s=1}^{p'} s^r y_s .$$

Conversely we can pass from the range $x=-p$ to $x=p$ to the range $x=1$ to $x=p'$, by means of the substitution:

$$p'=2p+1,$$

$$m_r=M_r+r\left(\frac{p'+1}{2}\right)M_{r-1}+\frac{r(r-1)}{2!}\left(\frac{p'+1}{2}\right)^2 M_{r-2}+\cdots\left(\frac{p'+1}{2}\right)^r M_0 .$$

Replacing $M_r$ and $m_r$ by $x^r$, it is clear that new polynomials $\psi_m(x)$ are obtained which belong to the range $x=1$ to $x=p'$. The polynomials may be explicitly evaluated from

$$\phi_m(x)=\phi_0+\phi_1 x+\cdot\cdot+\phi_m x^m$$

as follows:

$$\psi_m(x) = \phi_0' + \phi_1'(x-b) + \phi_2'(x^2 - 2bx + b^2) + \cdots \tag{16}$$
$$+ \phi_m'\left[x^m - mbx^{m-1} + m(m-1)b^2x^{m-2}/2 + \cdots\right],$$

where $b = (p'-1)/2$ and $\phi_r'$ denotes the value of $\phi_r$ after the substitution $p = (p'-1)/2$.

These polynomials can be proved by the method of section 3 to be orthogonal over the discrete range $x = 1$ to $x = p'$ and they are identifiable with the last lines of the formulas (3), (4), and (5) of the Annals paper previously cited where the $m_r$ are replaced by $x^r$.

Polynomials orthogonal over the discrete range $x = 0$ to $x = n-1$ were first obtained by Gram in the paper to which we previously referred and hence (16) may properly be called *the Gram polynomial of m th degree.*

The following explicit formula, in the notation of the present paper where the range is from $x = 1$ to $x = p$, was derived by Gram:

$$\Psi_m(x) = \left\{\frac{1}{(p-m-1)!}\right\}\left\{\frac{(p-1)!}{m!} - \frac{(m+1)(p-2)!(x-1)}{(m-1)!\cdot 1!^2}\right.$$
$$+ \frac{(m+1)(m+2)(p-3)!\,(x-1)(x-2)}{(m-2)!\,2!^2}$$
$$\left. - \frac{(m+1)(m+2)(m+3)(p-4)!(x-1)(x-2)(x-3)}{(m-3)!\,3!^2} + \cdots\right\}.$$

Since the coefficient of $x^m$ in $\psi_m(x)$ equals $\phi_m' = 4^m 1^2 \cdot 3^2 5^2 \cdots (2m-1)^2(2m+1)/(m!)^2 \cdot p(p^2-1)(p^2-4)\cdots(p^2-m^2)$, and since the coefficient of $x^m$ in Gram's definition is $(-1)^m(m+1)(m+2)\cdots\dfrac{(2m)}{(m!)^2}$, it is clear that the following equation holds between $\Psi_m(x)$ and $\psi_m(x)$:

$$\Psi_m(x)=\{(-1)^m p(p^2-1)(p^2-4)\cdots(p^2-m^2)(2m)!/4^m \cdot 1^2 \cdot 3^2 \cdot 5^2 \cdots$$

$$(2m-1)^2(2m+1)\,m!\}\,\psi_m(x),$$

$$=\{(-1)^m p(p^2-1)(p^2-4)\cdots(p^2-m^2)(m!)/(2m)!(2m+1)\}\,\psi_m(x).$$

By methods previously used it can be shown that,

$$\sum_{x=1}^{p} \psi_m^2(x)=\phi'(p).$$

The first four Gram polynomials are given below explicitly:[9]

$$\psi_0(x) = 1/p,$$

$$\psi_1(x)=\left[12/p(p^2-1)\right]\left[x-(p+1)/2\right],$$

$$\psi_2(x)=\left[180/p(p^2-1)(p^2-4)\right]\left[x^2-(p+1)x+(p+1)(p+2)/6\right],$$

$$\psi_3(x)=\left[2800/p(p^2-1)(p^2-4)(p^2-9)\right]\left[x^3-3(p+1)\frac{x^2}{2}+\right.$$

$$\left.\frac{1}{10}(6p^2+15p+11)x \quad -(p+1)(p+2)(p+3)/20\right].$$

6. *Tables and Numerical Application.* In tables 1 to 7 the numerical values of the coefficients of equations (3) to (9) have been tabulated for values of $p$ by half integers. For the case of the straight line the range of $p$ is from 0.5 to 100.0; for the parabola the range is from 1.0 to 100.0; for the cubic the range is from 1.5 to 50.0; for the other polynomials the range does not exceed $p = 25.0$ The tables have been computed to ten significant figures and have been checked by duplicate calculation.

[9]These polynomials are essentially the same as those employed by Jordan (loc. cit.) except that the summation in his work has taken over the numbers 0, 1, 2, . . . ., $n-1$. His polynomials are also expressed in terms of the Newton polynomials; $x(x-1)\cdots(x-n)$.

In illustration of the application of these tables to the numerical problem of polynomial approximation, we shall fit polynomials to the data employed by Karl Pearson in the same connection, his method being, however, the method of moments.[10] The data are from T. N. Thiele[11] and consist of a system of frequencies obtained from a game of patience (solitaire):

| Value of character | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 0 | 3 | 7 | 35 | 101 | 89 | 94 | 70 | 46 | 30 | 15 | 4 | 5 | 1 | 0 |
| Class marks $(x)$ | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Computing the moments, using the values of $x$ for this purpose, we obtain the following:

$$M_0 = 500; \quad M_1 = -570; \quad M_2 = 2728; \quad M_3 = -5508;$$

$$M_4 = 34108; \quad M_5 = -76380; \quad M_6 = 626188, \quad M_7 = -1419708.$$

These values are then substituted in equations (3) to (9) and the coefficients of the desired polynomials thus obtained. In illustration we shall give only the computations for the parabola.

From the value corresponding to $p=7$ in column $A'$ of table 1 we obtain,

$$a_1 = M_1 . (2)357\,1428\,571 = -2\;035\;714;$$

Similarly from the values corresponding to $p=7$ in columns $A$, $B$, and $C$ of table 2 we compute,

$$a_0 = M_0 . 151\,1312\,217 - M_2 \cdot (2)\,452\,4886\,878 = 63\,221\,719,$$

$$a_1 = -M_0 . (2)\,452\,4886\,878 + M_2 \cdot (3)\,242\,4046\,542 - 1.601\,1635$$

[10]On the Systematic Fitting of Curves to Observations and Measurements Biometrika, vol. 1 (1902), pp. 265-303; vol 2 (1903), pp. 1-27, in particular, p. 18.

[11]Forelaesninger over Almindelig Iagttagelseslaere, Copenhagen, (1889), p. 12

Proceeding in this manner the other coefficients are easily computed and we obtain the following seven polynomials of approximation:

$y = 33.333333 - 2.035714\,x,$

$y = 63.221719 - 2.035714x - 1.601163x^2,$

$y = 63.221719 - 9.924624x - 1.601163x^2 + .236195x^3,$

$y = 75.058367 - 9.924624x - 3.760517x^2 + .236195x^3 + .045666x^4,$

$y = 75.058367 - 20.405890x - 3.760517x^2 + 1.139793x^3 + .045666x^4$
$\quad - .014922x^5,$

$y = 73.950386 - 20.405890x - 3.320586x^2 + 1.139793x^3$
$\quad + .020890x^4 - .014922x^5 + .000338551x^6,$

$y = 73.950386 - 25.763034x - 3.320586x^2 + 2.070322x^3$
$\quad + .020890x^4 - .054118x^5 + .000338551x^6 + .00046071106x^7.$

The following table contains the values computed from these polynomials over the range from $x = -7$ to $x = 7$:

| $x$ | $y$ | $n=1$ | $n=2$ | $n=3$ | $n=4$ | $n=5$ | $n=6$ | $n=7$ |
|---|---|---|---|---|---|---|---|---|
| −7 | 0 | 47.583 | −.985 | −26.778 | −11.105 | 3.123 | 3.916 | .591 |
| −6 | 3 | 45.548 | 17.794 | 14.109 | 7.393 | −8.863 | −10.448 | −3.483 |
| −5 | 7 | 43.512 | 33.371 | 43.291 | 29.685 | 15.773 | 15.469 | 12.432 |
| −4 | 35 | 41.476 | 45.746 | 62.185 | 51.163 | 50.538 | 51.513 | 45.975 |
| −3 | 101 | 39.440 | 54.918 | 72.208 | 68.309 | 78.982 | 80.073 | 79.537 |
| −2 | 89 | 37.405 | 60.888 | 74.777 | 78.707 | 92.918 | 93.194 | 97.660 |
| −1 | 94 | 35.369 | 63.656 | 71.309 | 81.032 | 90.625 | 89.932 | 94.397 |
| 0 | 70 | 33.333 | 63.222 | 63.222 | 75.058 | 75.058 | 73.950 | 73.950 |
| 1 | 46 | 31.298 | 59.585 | 51.932 | 61.655 | 52.062 | 51.370 | 46.905 |
| 2 | 30 | 29.262 | 52.746 | 38.857 | 42.787 | 28.576 | 28.853 | 24.388 |
| 3 | 15 | 27.226 | 42.704 | 25.415 | 21.516 | 10.843 | 11.935 | 12.471 |
| 4 | 4 | 25.190 | 29.460 | 13.021 | 1.999 | 2.624 | 3.599 | 9.136 |
| 5 | 5 | 23.155 | 13.014 | 3.094 | −10.512 | 3.400 | 3.095 | 6.133 |
| 6 | 1 | 21.119 | −6.634 | −2.950 | −9.667 | 6.589 | 5.005 | −1.959 |
| 7 | 0 | 19.083 | −29.485 | −33.693 | 11.980 | −2.249 | −1.457 | .869 |

The approximation attained by these polynomials is exhibited

in the following charts, the odd order cases being given in figure 1 and the even order cases in figure 2.

In order to illustrate the case where the number of items is even we shall delete the last value from the series which we have just used. The table must then be arranged as follows:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Frequency | 0 | 3 | 7 | 35 | 101 | 89 | 94 | 70 |
| Class mark | -13/2 | -11/2 | -9/2 | -7/2 | -5/2 | -3/2 | -1/2 | 1/2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Frequency | 46 | 30 | 15 | 4 | 5 | 1 |
| Class mark | 3/2 | 5/2 | 7/2 | 9/2 | 11/2 | 13/2 |

The method will be sufficiently illustrated by means of the first, second, and fifth degree polynomials To compute these we first obtain the moments.

$$M_0 = 500, \quad M_1 = -320, \quad M_2 = 2283, \quad M_3 = -1781,$$
$$M_4 = 26930.25, \quad M_5 = -1632\ 5.$$

In order to evaluate the coefficients of the parabola we use the value corresponding to $\rho = 6.5$ in column $A'$ of table 1 and the values in columns $A$, $B$, and $C$ of table 2 as follows

$$a_1 = M_1\ (2)\ 439\ 5604\ 396 = -1.406593,$$

$$a_0 = M_0\ 162\ 1093\ 750 - M_2\ (2)\ 558\ 0357\ 143 = 68\ 314\ 734,$$

$$a_2 = -M_0\ (2)\ 558\ 0357\ 143 + M_2\ (3)\ 3434065\ 934 = -2\ 006\ 181$$

The other coefficients are similarly obtained and we thus derive the following approximating polynomials:

$$y = 35\ 714286 - 1.406593x,$$

$$y = 68\ 314734 - 1.406593x - 2\ 006181x^2;$$

$$y = 83\ 37250 - 16\ 63150x - 5\ 172034x^2 + 1\ 149221x^3 + 077082x^4$$
$$-017801x^5.$$

The approximating values obtained from the parabola and the quintic are recorded in the following table:

| $x$ | $y$ | $n=2$ | $n=5$ | $x$ | $y$ | $n=2$ | $n=5$ |
|---|---|---|---|---|---|---|---|
| -6.5 | 0 | - 7.304 | 1 492 | 0 5 | 70 | 67.110 | 73.912 |
| -5.5 | 3 | 15.364 | - 12 686 | 1.5 | 46 | 61 691 | 50 922 |
| -4.5 | 7 | 34 019 | 13.214 | 2.5 | 30 | 52 260 | 28.698 |
| -3.5 | 35 | 48.662 | 49.869 | 3.5 | 15 | 38.816 | 13.295 |
| -2.5 | 101 | 59.293 | 79.419 | 4.5 | 4 | 21.360 | 7.280 |
| -1.5 | 89 | 65.911 | 93 329 | 5.5 | 5 | -.108 | 7.592 |
| -0.5 | 94 | 68.516 | 90.257 | 6.5 | 1 | -25.589 | 3.407 |

In the article from which these data are taken, Karl Pearson compares the efficacy of polynomial curves with that of skew-frequency curves and shows the superiority of the latter in the present case. It is worth noting here, however, that the *least square* polynomials of the present paper give a measurably better fit than the *moment* polynomials employed by Pearson. The sum of the squares of the deviations from the data of the values obtained by means of the sixth degree parabola of Pearson is found to be 1402.31; the same sum for the sextic of the present paper is 1091.22. For the septic the sum of the square of the deviations is 926.32, which compares not too unfavorably with the sum 760.91 obtained from the skew-frequency curve.

H. T. Davis

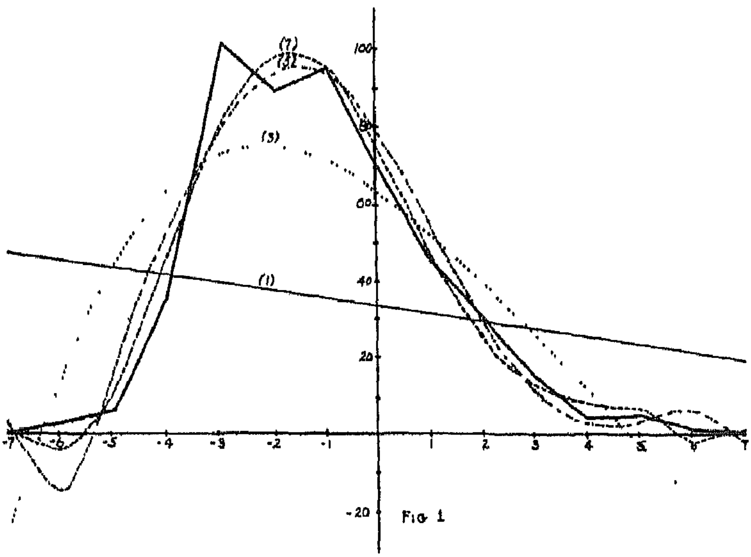

(7)
(3)
(1)
100
80
60
40
20
-20
-7
-6
-5
-4
-3
-2
-1
1
2
3
4
5
6
7

FIG 1

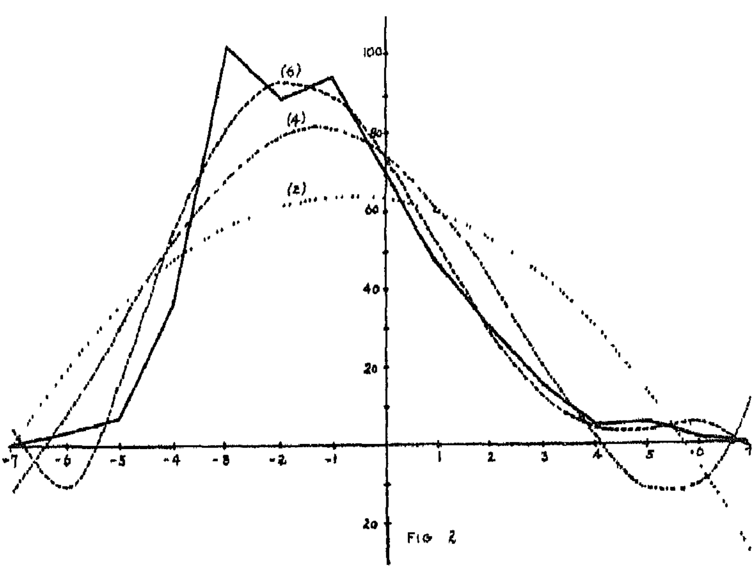

(6)
(4)
(2)
100
80
60
40
20
20
-7
-6
-5
-4
-3
-2
-1
1
2
3
4
5
6
7

FIG 2

TABLE I

(The numbers in parentheses denote the number of ciphers between the decimal point and the first significant figure.)

| $p$ | $A$ | $A'$ |
|---|---|---|
| 0 5 | .500 0000 000 | 2.000 0000 000 |
| 1.0 | .333 3333 333 | .500 0000 000 |
| 1.5 | .250 0000 000 | .200 0000 000 |
| 2.0 | .200 0000 000 | .100 0000 000 |
| 2.5 | 166 6666 667 | .(1)571 4285 714 |
| 3.0 | 142 8571 429 | .(1)357 1428 571 |
| 3.5 | .125 0000 000 | .(1)238 0952 381 |
| 4 0 | 111 1111 111 | .(1)166 6666 667 |
| 4.5 | .100 0000 000 | .(1)121 2121 212 |
| 5.0 | .(1)909 0909 091 | (2)909 0909 091 |
| 5.5 | .(1)833 3333 333 | .(2)699 3006 993 |
| 6 0 | .(1)769 2307 692 | (2)549 4505 495 |
| 6.5 | .(1)714 2857 143 | .(2)439 5604 396 |
| 7 0 | (1)666 6666 667 | (2)357 1428 571 |
| 7.5 | .(1)625 0000 000 | .(2)294 1176 471 |
| 8.0 | .(1)588 2352 941 | (2)245 0980 392 |
| 8 5 | .(1)555 5555 556 | .(2)206 3983 488 |
| 9.0 | .(1)526 3157 895 | .(2)175 4385 965 |
| 9.5 | .(1)500 0000 000 | .(2)150 3759 398 |
| 10.0 | .(1)476 1904 762 | .(2)129 8701 299 |
| 10.5 | .(1)454 5454 545 | .(2)112 9305 477 |
| 11.0 | .(1)434 7826 087 | (3)988 1422 925 |
| 11 5 | .(1)416 6666 667 | .(3)869 5652 174 |
| 12 0 | .(1)400 0000 000 | .(3)769 2307 692 |
| 12.5 | .(1)384 6153 846 | .(3)683 7606 838 |
| 13.0 | .(1)370 3703 704 | .(3)610 5006 105 |
| 13.5 | .(1)357 1428 571 | .(3)547 3453 749 |
| 14.0 | .(1)344 8275 862 | .(3)492 6108 374 |
| 14.5 | .(1)333 3333 333 | .(3)444 9388 209 |
| 15 0 | .(1)322 5806 452 | (3)403 2258 064 |
| 15.5 | .(1)312 5000 000 | .(3)366 5689 150 |
| 16.0 | .(1)303 0303 030 | .(3)334 2245 989 |
| 16 5 | .(1)294 1176 471 | .(3)305 5767 762 |
| 17 0 | .(1)285 7142 857 | (3)280 1120 448 |
| 17.5 | .(1)277 7777 778 | .(3)257 4002 574 |
| 18.0 | .(1)270 2702 703 | (3)237 0791 844 |
| 18.5 | .(1)263 1578 947 | .(3)218 8423 241 |
| 19.0 | .(1)256 4102 564 | .(3)202 4291 498 |
| 19.5 | .(1)250 0000 000 | .(3)187 6172 608 |
| 20.0 | .(1)243 9024 390 | .(3)174 2160 279 |
| 20 5 | (1)238 0952 381 | .(3)162 0614 213 |
| 21 0 | .(1)232 5581 395 | .(3)151 0117 789 |
| 21.5 | .(1)227 2727 273 | .(3)140 9443 270 |
| 22.0 | .(1)222 2222 222 | .(3)131 7523 057 |
| 22 5 | .(1)217 3913 043 | .(3)123 3425 840 |
| 23.0 | .(1)212 7659 574 | (3)115 6336 725 |
| 23.5 | .(1)208 3333 333 | .(3)108 5540 599 |
| 24 0 | .(1)204 0816 327 | (3)102 0408 163 |
| 24.5 | .(1)200 0000 000 | .(4)960 3841 537 |
| 25.0 | .(1)196 0784 314 | .(4)904 9773 756 |

TABLE I—(Continued)

| $\rho$ | $A$ | $A'$ |
|---|---|---|
| 25.5 | .(1)192 3076 923 | .(4)853 7522 411 |
| 26.0 | .(1)188 6792 453 | .(4)806 3215 610 |
| 26.5 | .(1)185 1851 852 | .(4)762 3403 850 |
| 27.0 | .(1)181 8181 818 | .(4)721 5007 215 |
| 27.5 | .(1)178 5714 286 | .(4)683 5269 993 |
| 28.0 | .(1)175 4385 965 | .(4)648 1721 545 |
| 28.5 | .(1)172 4137 931 | .(4)615 2142 484 |
| 29.0 | .(1)169 4915 254 | .(4)584 4535 359 |
| 29.5 | .(1)166 6666 667 | .(4)555 7099 194 |
| 30.0 | .(1)163 9344 262 | .(4)528 8207 298 |
| 30.5 | .(1)161 2903 226 | .(4)503 6387 903 |
| 31.0 | .(1)158 7301 587 | .(4)480 0307 220 |
| 31 5 | .(1)156 2500 000 | .(4)457 8754 579 |
| 32 0 | .(1)153 8461 538 | .(4)437 0629 371 |
| 32 5 | .(1)151 5151 515 | .(4)417 4929 548 |
| 33.0 | .(1)149 2537 313 | .(4)399 0741 480 |
| 33 5 | .(1)147 0588 235 | .(4)381 7230 981 |
| 34 0 | .(1)144 9275 362 | .(4)365 3635 367 |
| 34.5 | .(1)142 8571 429 | .(4)349 9256 408 |
| 35 0 | .(1)140 8450 704 | .(4)335 3454 058 |
| 35.5 | .(1)138 8888 889 | .(4)321 5640 877 |
| 36 0 | .(1)136 9863 014 | .(4)308 5277 058 |
| 36 5 | .(1)135 1351 351 | .(4)296 1865 976 |
| 37 0 | .(1)133 3333 333 | .(4)284 4950 213 |
| 37 5 | .(1)131 5789 474 | .(4)273 4107 997 |
| 38 0 | .(1)129 8701 299 | .(4)262 8949 997 |
| 38.5 | .(1)128 2051 282 | .(4)252 9116 453 |
| 39.0 | .(1)126 5822 785 | .(4)243 4274 586 |
| 39 5 | .(1)125 0000 000 | .(4)234 4116 268 |
| 40.0 | .(1)123 4567 901 | .(4)225 8355 917 |
| 40 5 | .(1)121 9512 195 | .(4)217 6728 595 |
| 41 0 | .(1)120 4819 277 | .(4)209 8988 288 |
| 41 5 | .(1)119 0476 190 | .(4)202 4906 348 |
| 42.0 | .(1)117 6470 588 | .(4)195 4270 080 |
| 42 5 | .(1)116 2790 698 | .(4)188 6881 457 |
| 43.0 | (1)114 9425 287 | .(4)182 2555 952 |
| 43 5 | .(1)113 6363 636 | .(4)176 1121 482 |
| 44 0 | .(1)112 3595 506 | .(4)170 2417 433 |
| 44 5 | .(1)111 1111 111 | .(4)164 6293 781 |
| 45 0 | .(1)109 8901 099 | .(4)159 2610 288 |
| 45 5 | .(1)108 6956 522 | .(4)154 1235 763 |
| 46.0 | .(1)107 5268 817 | .(4)149 2047 387 |
| 46 5 | .(1)106 3829 787 | .(4)144 4930 102 |
| 47.0 | .(1)105 2631 579 | .(4)139 9776 036 |
| 47 5 | .(1)104 1666 667 | .(4)135 6483 993 |
| 48.0 | .(1)103 0927 835 | .(4)131 4958 973 |
| 48.5 | .(1)102 0408 163 | .(4)127 5111 732 |
| 49.0 | .(1)101 0101 010 | .(4)123 6858 380 |
| 49.5 | .(1)100 0000 000 | .(4)120 0120 012 |
| 50 0 | .(2)990 0990 099 | .(4)116 4822 365 |

TABLE I—(Continued)

| $p$ | $A$ | $A'$ |
|---|---|---|
| 50.5 | .(2)980 3921 569 | .(4)113 0895 500 |
| 51.0 | .(2)970 8737 864 | .(4)109 8273 514 |
| 51.5 | .(2)961 5384 615 | .(4)106 6894 271 |
| 52 0 | .(2)952 3809 524 | .(4)103 6699 150 |
| 52 5 | .(2)943 3962 264 | .(4)100 7632 819 |
| 53.0 | .(2)934 5794 393 | .(5)979 6430 181 |
| 53.5 | .(2)925 9259 259 | .(5)952 6803 662 |
| 54 0 | .(2)917 4311 927 | .(5)926 6981 744 |
| 54 5 | .(2)909 0909 091 | .(5)901 6522 778 |
| 55.0 | .(2)900 9009 009 | .(5)877 5008 775 |
| 55.5 | .(2)892 8571 429 | .(5)854 2043 940 |
| 56.0 | .(2)884 9557 522 | .(5)831 7253 310 |
| 56 5 | .(2)877 1929 825 | .(5)810 0281 485 |
| 57 0 | .(2)869 5652 174 | .(5)789 0791 446 |
| 57.5 | .(2)862 0689 655 | .(5)768 8463 461 |
| 58 0 | .(2)854 7008 547 | .(5)749 2994 051 |
| 58.5 | .(2)847 4576 271 | .(5)730 4095 041 |
| 59.0 | .(2)840 3361 345 | .(5)712 1492 665 |
| 59.5 | .(2)833 3333 333 | .(5)694 4926 731 |
| 60.0 | .(2)826 4462 810 | .(5)677 4149 844 |
| 60.5 | .(2)819 6721 311 | .(5)660 8926 677 |
| 61.0 | (2)813 0081 301 | .(5)644 9033 290 |
| 61.5 | .(2)806 4516 129 | .(5)629 4256 491 |
| 62.0 | .(2)800 0000 000 | .(5)614 4393 241 |
| 62.5 | .(2)793 6507 937 | .(5)599 9250 094 |
| 63.0 | .(2)787 4015 748 | .(5)585 8642 670 |
| 63 5 | .(2)781 2500 000 | .(5)572 2395 166 |
| 64 0 | .(2)775 1937 984 | .(5)559 0339 893 |
| 64 5 | .(2)769 2307 692 | .(5)546 2316 842 |
| 65.0 | .(2)763 3587 786 | .(5)533 8173 277 |
| 65 5 | .(2)757 5757 576 | .(5)521 7763 354 |
| 66 0 | .(2)751 8796 992 | .(5)510 0947 756 |
| 66 5 | .(2)746 2686 567 | .(5)498 7593 361 |
| 67.0 | .(2)740 7407 407 | .(5)487 7572 920 |
| 67.5 | .(2)735 2941 176 | .(5)477 0764 754 |
| 68.0 | .(2)729 9270 073 | .(5)466 7052 476 |
| 68 5 | .(2)724 6376 812 | .(5)456 6324 725 |
| 69.0 | .(2)719 4244 604 | .(5)446 8474 910 |
| 69.5 | .(2)714 2857 143 | (5)437 3400 975 |
| 70.0 | .(2)709 2198 582 | .(5)428 1005 180 |
| 70.5 | .(2)704 2253 521 | .(5)419 1193 883 |
| 71.0 | .(2)699 3006 993 | .(5)410 3877 343 |
| 71 5 | .(2)694 4444 444 | .(5)401 8969 536 |
| 72.0 | .(2)689 6551 724 | .(5)393 6387 970 |
| 72.5 | .(2)684 9315 068 | .(5)385 6053 522 |
| 73.0 | .(2)680 2721 088 | .(5)377 7890 275 |
| 73 5 | .(2)675 6756 757 | .(5)370 1825 370 |
| 74.0 | .(2)671 1409 396 | .(5)362 7788 863 |
| 74 5 | .(2)666 6666 667 | .(5)355 5713 587 |
| 75.0 | .(2)662 2516 556 | .(5)348 5535 030 |

TABLE I—(Continued)

| $\rho$ | $A$ | $A'$ |
|---|---|---|
| 75.5 | .(2)657 8947 368 | .(5)341 7191 206 |
| 76.0 | .(2)653 5947 712 | .(5)335 0622 546 |
| 76.5 | .(2)649 3506 494 | .(5)328 5771 787 |
| 77.0 | .(2)645 1612 903 | .(5)322 2583 868 |
| 77.5 | .(2)641 0256 410 | .(5)316 1005 832 |
| 78.0 | .(2)636 9426 752 | .(5)310 0986 734 |
| 78.5 | .(2)632 9113 924 | .(5)304 2477 550 |
| 79.0 | .(2)628 9308 176 | .(5)298 5431 096 |
| 79.5 | .(2)625 0000 000 | .(5)292 9801 945 |
| 80.0 | .(2)621 1180 124 | .(5)287 5546 354 |
| 80.5 | .(2)617 2839 506 | .(5)282 2622 188 |
| 81.0 | .(2)613 4969 325 | .(5)277 0988 855 |
| 81.5 | (2)609 7560 976 | .(5)272 0607 240 |
| 82 0 | .(2)606 0606 061 | .(5)267 1439 639 |
| 82.5 | .(2)602 4096 386 | .(5)262 3449 705 |
| 83.0 | .(2)598 8023 952 | .(5)257 6602 389 |
| 83.5 | .(2)595 2380 952 | .(5)253 0863 885 |
| 84.0 | .(2)591 7159 763 | .(5)248 6201 581 |
| 84.5 | .(2)588 2352 941 | .(5)244 2584 010 |
| 85.0 | .(2)584 7953 216 | .(5)239 9980 800 |
| 85.5 | .(2)581 3953 488 | .(5)235 8362 636 |
| 86.0 | .(2)578 0346 821 | .(5)231 7701 211 |
| 86.5 | .(2)574 7126 437 | .(5)227 7969 190 |
| 87.0 | .(2)571 4285 714 | .(5)223 9140 170 |
| 87.5 | .(2)568 1818 182 | .(5)220 1188 642 |
| 88.0 | .(2)564 9717 514 | .(5)216 4089 957 |
| 88.5 | .(2)561 7977 528 | .(5)212 7820 293 |
| 89.0 | .(2)558 6592 179 | .(5)209 2350 621 |
| 89.5 | .(2)555 5555 556 | .(5)205 7676 677 |
| 90.0 | .(2)552 4861 878 | .(5)202 3758 930 |
| 90.5 | .(2)549 4505 495 | .(5)199 0582 554 |
| 91.0 | .(2)546 4480 874 | .(5)195 8127 404 |
| 91.5 | .(2)543 4782 609 | .(5)192 6373 986 |
| 92.0 | .(2)540 5405 405 | .(5)189 5303 438 |
| 92.5 | .(2)537 6344 086 | .(5)186 4897 501 |
| 93.0 | .(2)534 7593 583 | .(5)183 5138 498 |
| 93.5 | .(2)531 9148 936 | .(5)180 6009 315 |
| 94.0 | .(2)529 1005 291 | .(5)177 7493 378 |
| 94.5 | (2)526 3157 895 | .(5)174 9574 635 |
| 95.0 | .(2)523 5602 094 | .(5)172 2237 531 |
| 95.5 | .(2)520 8333 333 | .(5)169 5466 999 |
| 96.0 | .(2)518 1347 150 | .(5)166 9248 438 |
| 96.5 | .(2)515 4639 175 | .(5)164 3567 692 |
| 97.0 | .(2)512 8205 128 | .(5)161 8411 044 |
| 97.5 | .(2)510 2040 816 | .(5)159 3765 191 |
| 98.0 | .(2)507 6142 132 | .(5)156 9617 233 |
| 98.5 | .(2)505 0505 051 | .(5)154 5954 662 |
| 99.0 | .(2)502 5125 628 | .(5)152 2765 342 |
| 99.5 | .(2)500 0000 000 | .(5)150 0037 501 |
| 100.0 | .(2)497 5124 378 | .(5)147 7759 716 |

TABLE II

(The number in parentheses denote the number of ciphers between the decimal point and the first significant figure.)

| $\rho$ | $A$ | $B$ | $C$ |
|---|---|---|---|
| 1.0 | 1.000 0000 000 | − 1.000 0000 000 | 1.500 0000 000 |
| 1.5 | .640 6250 000 | − .312 5000 000 | .250 0000 000 |
| 2 0 | .485 7142 857 | − 142 8571 429 | .(1)714 2857 143 |
| 2.5 | .394 5312 500 | −.(1)781 2500 000 | .(1)267 8571 429 |
| 3.0 | .333 3333 333 | −.(1)476 1904 762 | .(1)119 0476 190 |
| 3 5 | .289 0625 000 | −.(1)312 5000 000 | .(2)595 2380 952 |
| 4.0 | .255 4112 554 | −.(1)216 4502 165 | .(2)324 6753 247 |
| 4.5 | .228 9062 500 | −.(1)156 2500 000 | .(2)189 3939 394 |
| 5 0 | .207 4592 075 | −.(1)116 5501 166 | .(2)116 5501 166 |
| 5.5 | .189 7321 429 | −.(2)892 8571 429 | .(3)749 2507 493 |
| 6.0 | .174 8251 748 | −.(2)699 3006 993 | .(3)499 5004 995 |
| 6.5 | .162 1093 750 | −.(2)558 0357 143 | .(3)343 4065 934 |
| 7.0 | .151 1312 217 | −.(2)452 4886 878 | .(3)242 4046 542 |
| 7.5 | .141 5550 595 | −.(2)372 0238 095 | .(3)175 0700 280 |
| 8.0 | .133 1269 350 | −.(2)309 5975 232 | (3)128 9989 680 |
| 8 5 | .125 6510 417 | −.(2)260 4166 667 | .(4)967 4922 601 |
| 9 0 | .118 9739 054 | −.(2)221 1410 880 | .(4)737 1369 601 |
| 9.5 | .112 9734 848 | −.(2)189 3939 394 | .(4)569 6058 328 |
| 10.0 | 107 5514 874 | −.(2)163 4521 085 | .(4)445 7784 778 |
| 10 5 | .102 6278 409 | −.(2)142 0454 545 | .(4)352 9079 616 |
| 11.0 | .(1)981 3664 596 | −.(2)124 2236 025 | .(4)282 3263 693 |
| 11.5 | .(1)940 2316 434 | −.(2)109 2657 343 | .(4)228 0328 367 |
| 12 0 | .(1)902 4154 589 | −.(3)966 1835 749 | .(4)185 8045 336 |
| 12 5 | .(1)867 5309 066 | −.(3)858 5164 835 | (4)152 6251 526 |
| 13.0 | (1)835 2490 421 | −.(3)766 2835 249 | .(4)126 3104 711 |
| 13.5 | .(1)805 2884 615 | −.(3)686 8131 868 | .(4)105 2587 260 |
| 14 0 | .(1)777 4069 954 | −.(3)617 9705 846 | .(5)882 8151 209 |
| 14.5 | .(1)751 3950 893 | −.(3)558 0357 143 | .(5)744 8752 582 |
| 15.0 | .(1)727 0704 824 | − (3)505 6122 965 | .(5)632 0153 706 |
| 15.5 | .(1)704 2738 971 | −.(3)459 5588 235 | .(5)539 0719 338 |
| 16.0 | .(1)682 8655 216 | − (3)418 9359 028 | .(5)462 0616 575 |
| 16.5 | .(1)662 7221 201 | −.(3)382 9656 863 | .(5)397 8864 273 |
| 17.0 | .(1)643 7346 437 | −.(3)351 0003 510 | .(5)344 1179 912 |
| 17.5 | .(1)625 8062 436 | −.(3)322 4974 200 | .(5)298 8393 081 |
| 18.0 | .(1)608 8506 089 | − (3)297 0002 970 | .(5)260 5265 763 |
| 18 5 | .(1)592 7905 702 | −.(3)274 1228 070 | .(5)227 9607 543 |
| 19 0 | .(1)577 5569 190 | −.(3)253 5368 389 | (5)200 1606 623 |
| 19 5 | .(1)563 0874 060 | −.(3)234 9624 060 | .(5)176 3320 120 |
| 20.0 | .(1)549 3258 868 | − (3)218 1596 056 | .(5)155 8282 897 |
| 20.5 | .(1)536 2215 909 | − (3)202 9220 779 | .(5)138 1205 295 |
| 21.0 | .(1)523 7284 931 | −.(3)189 0716 582 | .(5)122 7738 040 |
| 21.5 | .(1)511 8047 713 | −.(3)176 4539 808 | .(5)109 4288 253 |
| 22.0 | .(1)500 4123 371 | −.(3)164 9348 507 | .(6)977 8746 091 |
| 22 5 | .(1)489 5164 279 | −.(3)154 3972 332 | .(6)876 0126 706 |
| 23 0 | .(1)479 0852 511 | −.(3)144 7387 466 | .(6)785 2017 048 |
| 23.5 | (1)469 0896 739 | −.(3)135 8695 652 | .(6)707 9612 604 |
| 24.0 | .(1)459 5029 501 | −.(3)127 7106 587 | .(6)638 5532 937 |
| 24.5 | .(1)450 3004 808 | −.(3)120 1923 077 | .(6)577 1539 385 |
| 25.0 | .(1)441 4596 027 | − (3)113 2528 483 | .(6)522 7054 537 |

TABLE II—(Continued)

| $\rho$ | $A$ | $B$ | $C$ |
|---|---|---|---|
| 25.5 | .(1)432 9594 017 | −.(3)106 8376 068 | .(6)474 3068 006 |
| 26.0 | .(1)424 7805 469 | −.(3)100 8979 921 | .(6)431 1880 006 |
| 26.5 | .(1)416 9051 435 | −.(4)953 9072 039 | .(6)392 6890 719 |
| 27.0 | .(1)409 3166 020 | −.(4)902 7715 085 | .(6)358 2426 621 |
| 27.5 | .(1)401 9995 211 | −.(4)855 2271 483 | .(6)327 3596 740 |
| 28 0 | .(1)394 9395 832 | −.(4)810 9642 365 | .(6)299 6173 287 |
| 28.5 | .(1)388 1234 606 | −.(4)769 7044 335 | .(6)274 6492 180 |
| 29.0 | .(1)381 5387 315 | − (4)731 1972 624 | .(6)252 1369 870 |
| 29.5 | .(1)375 1738 042 | −.(4)695 2169 077 | .(6)231 8033 590 |
| 30.0 | .(1)369 0178 489 | −.(4)661 5594 279 | .(6)213 4062 671 |
| 30.5 | .(1)363 0607 359 | −.(4)630 0403 226 | .(6)196 7339 024 |
| 31.0 | .(1)357 2929 802 | −.(4)600 4924 038 | .(6)181 6005 253 |
| 31.5 | .(1)351 7056 910 | −.(4)572 7639 296 | .(6)167 8429 098 |
| 32.0 | .(1)346 2905 254 | −.(4)546 7169 646 | .(6)155 3173 195 |
| 32.5 | .(1)341 0396 474 | −.(4)522 2259 358 | .(6)143 8969 284 |
| 33.0 | .(1)335 9456 896 | − (4)499 1763 590 | .(6)133 4696 147 |
| 33 5 | .(1)331 0017 189 | −.(4)477 4637 128 | .(6)123 9360 708 |
| 34.0 | .(1)326 2012 046 | −.(4)456 9924 414 | .(6)115 2081 785 |
| 34.5 | .(1)321 5379 902 | −.(4)437 6750 700 | .(6)107 2076 105 |
| 35.0 | .(1)317 0062 663 | −.(4)419 4314 188 | .(7)998 6462 352 |
| 35.5 | .(1)312 6005 470 | −.(4)402 1879 022 | .(7)931 1701 382 |
| 36.0 | .(1)308 3156 473 | −.(4)385 8769 053 | .(7)869 0921 290 |
| 36.5 | .(1)304 1466 631 | −.(4)370 4362 257 | .(7)811 9150 152 |
| 37.0 | .(1)300 0889 521 | −.(4)355 8085 750 | .(7)759 1932 610 |
| 37 5 | .(1)296 1381 169 | −.(4)341 9411 314 | .(7)710 5270 263 |
| 38.0 | .(1)292 2899 885 | −.(4)328 7851 389 | .(7)665 5569 614 |
| 38.5 | .(1)288 5406 123 | −.(4)316 2955 466 | .(7)623 9596 513 |
| 39.0 | .(1)284 8862 343 | −.(4)304 4306 842 | .(7)585 4436 234 |
| 39.5 | .(1)281 3232 880 | −.(4)293 1519 700 | (7)550 4855 968 |
| 40 0 | .(1)277 8483 837 | −.(4)282 4236 468 | .(7)516 6286 221 |
| 40.5 | .(1)274 4582 970 | −.(4)272 2125 436 | .(7)485 8769 184 |
| 41.0 | .(1)271 1499 593 | −.(4)262 4878 599 | .(7)457 2959 232 |
| 41 5 | .(1)267 9204 481 | −.(4)253 2209 708 | .(7)430 7052 407 |
| 42.0 | .(1)264 7669 787 | −.(4)244 3852 489 | .(7)405 9555 630 |
| 42.5 | .(1)261 6868 960 | −.(4)235 9559 046 | (7)382 8899 060 |
| 43.0 | .(1)258 6776 671 | −.(4)227 9098 389 | (7)361 3792 371 |
| 43 5 | .(1)255 7368 746 | −.(4)220 2255 109 | .(7)341 3026 128 |
| 44.0 | .(1)252 8622 095 | −.(4)212 8828 165 | .(7)322 5542 998 |
| 44.5 | .(1)250 0514 657 | −.(4)205 8629 776 | .(7)305 0198 458 |
| 45.0 | .(1)247 3025 344 | −.(4)199 1484 413 | .(7)288 6209 294 |
| 45.5 | .(1)244 6133 981 | −.(4)192 7227 875 | .(7)273 2687 523 |
| 46.0 | .(1)241 9821 265 | −.(4)186 5706 450 | .(7)258 8861 864 |
| 46.5 | .(1)239 4068 715 | −.(4)180 6776 133 | .(7)245 4025 308 |
| 47.0 | .(1)236 8858 628 | −.(4)175 0301 927 | .(7)232 7529 158 |
| 47.5 | .(1)234 4174 039 | −.(4)169 6157 186 | .(7)220 8777 671 |
| 48.0 | .(1)231 9998 685 | −.(4)164 4223 022 | .(7)209 7223 243 |
| 48.5 | .(1)229 6316 964 | −.(4)159 4387 755 | .(7)199 2362 081 |
| 49.0 | .(1)227 3113 909 | −.(4)154 6546 407 | .(7)189 3730 295 |
| 49.5 | .(1)225 0375 150 | −.(4)150 0600 240 | .(7)180 0900 478 |
| 50.0 | .(1)222 8086 886 | −.(4)145 6456 325 | (7)171 3478 030 |

TABLE II—(Continued)

| $\rho$ | $A$ | $B$ | $C$ |
|---|---|---|---|
| 50.5 | .(1)220 6235 860 | –.(4)141 4027 149 | .(7)163 1099 278 |
| 51.0 | .(1)218 4809 327 | –.(4)137 3230 250 | .(7)155 3427 884 |
| 51.5 | .(1)216 3795 035 | –.(4)133 3987 877 | .(7)148 0152 984 |
| 52.0 | .(1)214 3181 200 | –.(4)129 6226 684 | .(7)141 0986 957 |
| 52.5 | .(1)212 2956 479 | –.(4)125 9877 439 | .(7)134 5663 486 |
| 53.0 | .(1)210 3109 957 | –.(4)122 4874 757 | (7)128 3935 804 |
| 53.5 | .(1)208 3631 123 | –.(4)119 1156 852 | .(7)122 5575 085 |
| 54.0 | .(1)206 4509 850 | –.(4)115 8665 310 | .(7)117 0369 000 |
| 54 5 | .(1)204 5736 382 | –.(4)112 7344 877 | .(7)111 8120 384 |
| 55.0 | .(1)202 7301 313 | –.(4)109 7143 258 | .(7)106 8646 031 |
| 55.5 | .(1)200 9195 574 | –.(4)106 8010 936 | .(7)102 1775 591 |
| 56 0 | .(1)199 1410 418 | –.(4)103 9901 001 | .(8)977 3505 652 |
| 56.5 | .(1)197 3937 403 | –.(4)101 2768 991 | .(8)935 2233 857 |
| 57.0 | .(1)195 6768 382 | –.(5)986 5727 449 | (8)895 2565 744 |
| 57.5 | (1)193 9895 490 | –.(5)961 2722 631 | .(8)857 3219 737 |
| 58 0 | .(1)192 3311 130 | –.(5)936 8295 813 | .(8)821 3000 421 |
| 58.5 | .(1)190 7007 963 | – (5)913 2086 499 | (8)787 0792 070 |
| 59.0 | .(1)189 0978 896 | –.(5)890 3752 219 | .(8)754 5552 728 |
| 59.5 | .(1)187 5217 074 | –.(5)868 2967 491 | .(8)723 6308 764 |
| 60.0 | .(1)185 9715 868 | –.(5)846 9422 843 | .(8)694 2149 871 |
| 60.5 | .(1)184 4468 866 | – (5)826 2823 903 | .(8)666 2224 473 |
| 61.0 | .(1)182 9469 865 | –.(5)806 2890 546 | .(8)639 5735 494 |
| 61.5 | .(1)181 4712 863 | –.(5)786 9356 098 | .(8)614 1936 467 |
| 62.0 | .(1)180 0192 049 | –.(5)768 1966 583 | .(8)590 0127 944 |
| 62.5 | .(1)178 5901 798 | –.(5)750 0480 031 | .(8)566 9654 196 |
| 63.0 | .(1)177 1836 660 | –.(5)732 4665 812 | .(8)544 9900 158 |
| 63 5 | .(1)175 7991 358 | –.(5)715 4304 029 | .(8)524 0288 613 |
| 64.0 | .(1)174 4360 776 | –.(5)698 9184 935 | (8)504 0277 597 |
| 64 5 | .(1)173 0939 958 | –.(5)682 9108 392 | (8)484 9357 992 |
| 65.0 | .(1)171 7724 099 | –.(5)667 3883 359 | .(8)466 7051 300 |
| 65.5 | .(1)170 4708 538 | –.(5)652 3327 419 | .(8)449 2907 595 |
| 66.0 | .(1)169 1888 755 | –.(5)637 7266 321 | .(8)432 6503 610 |
| 66.5 | .(1)167 9260 366 | –.(5)623 5533 562 | .(8)416 7440 977 |
| 67.0 | .(1)166 6819 116 | –.(5)609 7969 986 | .(8)401 5344 591 |
| 67.5 | .(1)165 4560 875 | –.(5)596 4423 407 | .(8)386 9861 091 |
| 68.0 | .(1)164 2481 635 | –.(5)583 4748 260 | .(8)373 0657 455 |
| 68.5 | (1)163 0577 503 | –.(5)570 8805 261 | .(8)359 7419 689 |
| 69 0 | (1)161 8844 697 | –.(5)558 6461 100 | .(8)346 9851 615 |
| 69.5 | .(1)160 7279 547 | –.(5)546 7588 138 | .(8)334 7673 741 |
| 70.0 | .(1)159 5878 482 | –.(5)535 2064 131 | .(8)323 0622 212 |
| 70.5 | .(1)158 4638 037 | –.(5)523 9771 965 | .(8)311 8447 829 |
| 71.0 | .(1)157 3554 838 | –.(5)513 0599 408 | .(8)301 0915 145 |
| 71.5 | .(1)156 2625 611 | –.(5)502 4438 871 | .(8)290 7801 613 |
| 72 0 | .(1)155 1847 168 | –.(5)492 1187 187 | .(8)280 8896 796 |
| 72 5 | .(1)154 1216 409 | –.(5)482 0745 403 | .(8)271 4001 634 |
| 73.0 | .(1)153 0730 320 | –.(5)472 3018 576 | .(8)262 2927 754 |
| 73 5 | .(1)152 0385 968 | –.(5)462 7915 587 | .(8)253 5496 829 |
| 74.0 | .(1)151 0180 498 | –.(5)453 5348 963 | .(8)245 1539 980 |
| 74.5 | .(1)150 0111 131 | – (5)444 5234 708 | .(8)237 0897 218 |
| 75.0 | .(1)149 0175 162 | –.(5)435 7492 141 | .(8)229 3416 916 |

TABLE II—(Continued)

| $p$ | $A$ | $B$ | $C$ |
|---|---|---|---|
| 75.5 | .(1)148 0369 959 | −.(5)427 2043 746 | .(8)221 8955 328 |
| 76.0 | .(1)147 0692 956 | −.(5)418 8815 026 | .(8)214 7376 124 |
| 76.5 | .(1)146 1141 654 | −.(5)410 7734 371 | .(8)207 8549 966 |
| 77.0 | .(1)145 1713 622 | −.(5)402 8732 923 | .(8)201 2354 107 |
| 77.5 | .(1)144 2406 486 | −.(5)395 1744 458 | .(8)194 8672 015 |
| 78.0 | .(1)143 3217 937 | −.(5)387 6705 266 | .(8)188 7393 021 |
| 78.5 | .(1)142 4145 722 | −.(5)380 3554 041 | .(8)182 8411 989 |
| 79.0 | .(1)141 5187 645 | −.(5)373 2231 778 | .(8)177 1629 008 |
| 79.5 | .(1)140 6341 567 | −.(5)366 2681 669 | .(8)171 6949 101 |
| 80 0 | .(1)139 7605 399 | −.(5)359 4849 013 | (8)166 4281 950 |
| 80.5 | .(1)138 8977 106 | −.(5)352 8681 120 | .(8)161 3541 647 |
| 81.0 | .(1)138 0454 701 | −.(5)346 4127 230 | .(8)156 4646 446 |
| 81.5 | .(1)137 2036 248 | −.(5)340 1138 429 | .(8)151 7518 541 |
| 82.0 | .(1)136 3719 855 | −.(5)333 9667 569 | .(8)147 2083 854 |
| 82.5 | .(1)135 5503 678 | −.(5)327 9669 199 | .(8)142 8271 834 |
| 83.0 | .(1)134 7385 917 | −.(5)322 1099 490 | .(8)138 6015 271 |
| 83.5 | .(1)133 9364 812 | −.(5)316 3916 169 | .(8)134 5250 116 |
| 84.0 | .(1)133 1438 649 | −.(5)310 8078 455 | .(8)130 5915 317 |
| 84.5 | .(1)132 3605 750 | −.(5)305 3547 000 | .(8)126 7952 663 |
| 85.0 | .(1)131 5864 481 | −.(5)300 0283 827 | .(8)123 1306 632 |
| 85.5 | .(1)130 8213 241 | −.(5)294 8252 276 | .(8)119 5924 258 |
| 86.0 | .(1)130 0650 470 | −.(5)289 7416 953 | .(8)116 1754 993 |
| 86.5 | .(1)129 3174 642 | −.(5)284 7743 676 | .(8)112 8750 590 |
| 87.0 | .(1)128 5784 266 | −.(5)279 9199 429 | .(8)109 6864 980 |
| 87.5 | .(1)127 8477 885 | −.(5)275 1752 316 | .(8)106 6054 166 |
| 88.0 | .(1)127 1254 075 | −.(5)270 5371 515 | .(8)103 6276 117 |
| 88.5 | .(1)126 4111 445 | −.(5)266 0027 239 | .(8)100 7490 669 |
| 89.0 | .(1)125 7048 632 | −.(5)261 5690 691 | .(9)979 6594 350 |
| 89.5 | .(1)125 0064 308 | −.(5)257 2334 033 | .(9)952 7457 143 |
| 90.0 | .(1)124 3157 171 | −.(5)252 9930 341 | .(9)926 7144 106 |
| 90.5 | .(1)123 6325 948 | −.(5)248 8453 575 | .(9)901 5319 538 |
| 91.0 | .(1)122 9569 394 | −.(5)244 7878 546 | .(9)877 1662 253 |
| 91.5 | .(1)122 2886 291 | −.(5)240 8180 879 | .(9)853 5864 881 |
| 92.0 | .(1)121 6275 450 | −.(5)236 9336 988 | .(9)830 7633 199 |
| 92.5 | .(1)120 9735 702 | −.(5)233 1324 043 | .(9)808 6685 508 |
| 93.0 | .(1)120 3265 909 | −.(5)229 4119 941 | .(9)787 2752 029 |
| 93.5 | .(1)119 6864 953 | −.(5)225 7703 284 | .(9)766 5574 344 |
| 94.0 | .(1)119 0531 742 | −.(5)222 2053 346 | .(9)746 4904 858 |
| 94.5 | .(1)118 4265 205 | −.(5)218 7150 056 | .(9)727 0506 294 |
| 95.0 | .(1)117 8064 296 | −.(5)215 2973 968 | .(9)708 2151 209 |
| 95.0 | .(1)117 1927 988 | −.(5)211 9506 240 | .(9)689 9621 539 |
| 96.0 | .(1)116 5855 277 | −.(5)208 6728 615 | .(9)672 2708 166 |
| 96.5 | .(1)115 9845 180 | −.(5)205 4623 396 | .(9)655 1210 509 |
| 97.0 | .(1)115 3896 733 | −.(5)202 3173 428 | .(9)638 4936 130 |
| 97.5 | .(1)114 8008 993 | −.(5)199 2362 081 | .(9)622 3700 369 |
| 98.0 | .(1)114 2181 034 | −.(5)196 2173 225 | .(9)606 7325 988 |
| 98.5 | .(1)113 6411 951 | −.(5)193 2591 218 | .(9)591 5642 838 |
| 99.0 | .(1)113 0700 856 | −.(5)190 3600 889 | .(9)576 8487 544 |
| 99.5 | .(1)112 5046 880 | −.(5)187 5187 519 | .(9)562 5703 199 |
| 100.0 | .(1)111 9449 168 | −.(5)184 7336 824 | .(9)548 7139 081 |

TABLE III

(The numbers in parentheses denote the number of ciphers between the decimal point and the first significant figure.)

| $\rho$ | $A'$ | $B'$ | $C'$ |
|---|---|---|---|
| 1.5 | 2.534 7222 222 | − 1.138 8888 889 | .555 5555 556 |
| 2.0 | .902 7777 778 | −.236 1111 111 | .(1)694 4444 444 |
| 2.5 | .450 6999 559 | −.(1)779 3209 877 | .(1)154 3209 877 |
| 3.0 | .262 5661 376 | −.(1)324 0740 741 | .(2)462 9629 630 |
| 3.5 | .167 8541 366 | −.(1)155 7239 057 | .(2)168 3501 684 |
| 4.0 | .114 3378 227 | −.(2)827 7216 611 | .(3)701 4590 348 |
| 4.5 | .(1)816 0531 598 | −.(2)474 2942 243 | .(3)323 7503 238 |
| 5.0 | .(1)603 7943 538 | −.(2)288 1377 881 | .(3)161 8751 619 |
| 5.5 | .(1)459 7794 181 | −.(2)183 4585 167 | .(4)863 3341 967 |
| 6.0 | .(1)358 4609 835 | −.(2)121 4063 714 | .(4)485 6254 856 |
| 6.5 | .(1)285 0269 624 | −.(3)829 8482 563 | (4)285 6620 504 |
| 7.0 | .(1)230 4589 927 | −.(3)583 0679 850 | .(4)174 5712 530 |
| 7.5 | .(1)189 0399 941 | −.(3)419 5222 848 | .(4)110 2555 282 |
| 8.0 | .(1)157 0204 105 | −.(3)308 1642 014 | .(5)716 6609 334 |
| 8.5 | .(1)131 8685 978 | −.(3)230 5259 336 | .(5)477 7739 556 |
| 9.0 | .(1)111 8316 811 | −.(3)175 2561 737 | .(5)325 7549 697 |
| 9.5 | .(2)956 6897 397 | −.(3)135 1741 492 | .(5)226 6121 528 |
| 10.0 | .(2)824 8507 009 | −.(3)105 6201 476 | .(5)160 5169 416 |
| 10.5 | .(2)716 2246 443 | −.(4)835 0091 302 | .(5)115 5721 980 |
| 11.0 | .(2)625 9079 085 | −.(4)667 2071 889 | .(6)844 5660 620 |
| 11.5 | .(2)550 1917 791 | −.(4)538 3326 639 | .(6)625 6044 903 |
| 12.0 | .(2)486 2354 500 | −.(4)438 2359 455 | .(6)469 2033 678 |
| 12.5 | .(2)431 8375 890 | −.(4)359 6848 299 | .(6)355 9473 824 |
| 13.0 | .(2)385 2742 263 | −.(4)297 4533 626 | .(6)272 8929 932 |
| 13.5 | .(2)345 1820 327 | −.(4)247 7164 138 | .(6)211 2719 947 |
| 14.0 | .(2)310 4731 565 | −.(4)207 6407 573 | .(6)165 0562 459 |
| 14.5 | .(2)280 2723 203 | −.(4)175 1046 701 | .(6)130 0443 149 |
| 15.0 | .(2)253 8698 342 | −.(4)148 5029 580 | .(6)103 2704 854 |
| 15.5 | .(2)230 6861 266 | −.(4)126 6096 151 | .(7)826 1638 832 |
| 16.0 | .(2)210 2447 093 | − (4)108 4799 077 | .(7)665 5209 059 |
| 16.5 | .(2)192 1513 833 | −.(5)933 7977 791 | .(7)539 6115 453 |
| 17.0 | .(2)176 0781 076 | −.(5)807 3440 736 | .(7)440 2094 185 |
| 17.5 | .(2)161 7503 875 | −.(5)700 9036 937 | .(7)361 1974 716 |
| 18.0 | (2)148 9373 393 | −.(5)610 8752 239 | .(7)297 9879 141 |
| 18.5 | .(2)137 4438 092 | −.(5)534 3795 459 | .(7)247 1119 288 |
| 19.0 | .(2)127 1040 798 | −.(5)469 1008 114 | .(7)205 9266 073 |
| 19.5 | .(2)117 7768 137 | −.(5)413 1653 981 | .(7)172 4036 712 |
| 20.0 | .(2)109 3409 664 | −.(5)365 0491 008 | (7)144 9758 144 |
| 20.5 | .(2)101 6924 641 | −.(5)323 5054 757 | .(7)122 4240 211 |
| 21.0 | (3)947 4149 003 | −.(5)287 5101 521 | (7)103 7942 787 |
| 21.5 | .(3)884 1025 513 | −.(5)256 2172 813 | (8)883 3555 638 |
| 22.0 | .(3)826 3115 900 | −.(5)228 9252 750 | .(8)754 5328 774 |
| 22.5 | .(3)773 4526 546 | −.(5)205 0496 990 | .(8)646 7424 663 |
| 23.0 | .(3)725 0103 342 | −.(5)184 1017 105 | .(8)556 1985 210 |
| 23.5 | .(3)680 5325 601 | −.(5)165 6708 183 | (8)479 8575 476 |
| 24.0 | .(3)639 6217 018 | −.(5)149 4110 299 | .(8)415 2613 392 |
| 24.5 | .(3)601 9270 658 | −.(5)135 0296 678 | .(8)360 4155 020 |
| 25.0 | .(3)567 1385 543 | −.(5)122 2783 008 | .(8)313 6949 739 |

TABLE III—(Continued)

| $\rho$ | $A'$ | $B'$ | $C'$ |
|---|---|---|---|
| 25.5 | (3)534 9812 832 | -.(5)110 9453 570 | .(8)273 7701 591 |
| 26.0 | .(3)505 2110 028 | -.(5)100 8500 824 | .(8)239 5488 892 |
| 26.5 | (3)477 6101 850 | -.(6)918 3758 072 | .(8)210 1306 046 |
| 27.0 | .(3)451 9846 731 | -.(6)837 7472 451 | .(8)184 7700 144 |
| 27.5 | .(3)428 1608 021 | -.(6)765 4677 208 | .(8)162 8481 482 |
| 28.0 | .(3)405 9829 189 | -.(6)700 5455 924 | .(8)143 8491 976 |
| 28.5 | .(3)385 3112 389 | -.(6)642 1215 945 | .(8)127 3419 126 |
| 29.0 | .(3)366 0199 892 | -.(6)589 4492 824 | .(8)112 9645 999 |
| 29.5 | .(3)347 9957 969 | -.(6)541 8786 342 | .(8)100 4129 777 |
| 30.0 | .(3)331 1362 854 | -.(6)498 8422 596 | .(9)894 3030 827 |
| 30.5 | .(3)315 3488 491 | -.(6)459 8437 659 | .(9)797 9935 200 |
| 31.0 | .(3)300 5495 828 | -.(6)424 4479 170 | .(9)713 3578 436 |
| 31.5 | .(3)286 6623 418 | -.(6)392 2722 841 | .(9)638 8279 197 |
| 32.0 | .(3)273 6179 176 | -.(6)362 9801 450 | .(9)573 0662 221 |
| 32.5 | .(3)261 3533 111 | -.(6)336 2744 286 | .(9)514 9290 691 |
| 33.0 | .(3)249 8110 923 | -.(6)311 8925 371 | .(9)463 4361 622 |
| 33.5 | .(3)238 9388 342 | -.(6)289 6019 105 | .(9)417 7452 730 |
| 34.0 | .(3)228 6886 114 | -.(6)269 1962 143 | .(9)377 1311 492 |
| 34.5 | .(3)219 0165 553 | -.(6)250 4920 591 | .(9)340 9678 883 |
| 35.0 | .(3)209 8824 591 | -.(6)233 3261 690 | .(9)308 7141 691 |
| 35.5 | .(3)201 2494 260 | -.(6)217 5529 331 | .(9)279 9008 467 |
| 36.0 | .(3)193 0835 554 | -.(6)203 0422 839 | .(9)254 1205 055 |
| 36.5 | .(3)185 3536 630 | -.(6)189 6778 555 | .(9)231 0186 414 |
| 37.0 | .(3)178 0310 300 | -.(6)177 3553 804 | .(9)210 2861 992 |
| 37.5 | .(3)171 0891 785 | -.(6)165 9829 799 | .(9)191 6532 449 |
| 38.0 | .(3)164 5036 705 | -.(6)155 4715 079 | .(9)174 8835 859 |
| 38.5 | .(3)158 2519 263 | -.(6)145 7503 555 | .(9)159 7701 896 |
| 39.0 | .(3)152 3130 623 | -.(6)136 7496 434 | .(9)146 1312 710 |
| 39.5 | .(3)146 6677 431 | -.(6)128 4078 366 | .(9)133 8069 469 |
| 40.0 | .(3)141 2980 508 | -.(6)120 6693 349 | .(9)122 6563 680 |
| 40.5 | .(3)136 1873 638 | -.(6)113 4838 362 | .(9)112 5552 554 |
| 41.0 | .(3)131 3202 493 | -.(6)106 8057 759 | .(9)103 3937 811 |
| 41.5 | .(3)126 6823 653 | -.(6)100 5938 300 | .(10)950 7474 123 |
| 42.0 | .(3)122 2603 712 | -.(7)948 1047 667 | .(10)875 1197 773 |
| 42.5 | .(3)118 0418 479 | -.(7)894 2160 708 | .(10)806 2901 319 |
| 43.0 | .(3)114 0152 237 | -.(7)843 9617 986 | .(10)743 5786 772 |
| 43.5 | .(3)110 1697 079 | -.(7)797 0591 436 | .(10)686 3803 174 |
| 44.0 | .(3)106 4952 300 | -.(7)753 2501 737 | .(10)634 1557 280 |
| 44.5 | .(3)102 9823 841 | -.(7)712 2993 971 | .(10)586 4235 765 |
| 45.0 | .(4)996 2237 840 | -.(7)673 9915 889 | .(10)542 7537 357 |
| 45.5 | .(4)964 0698 884 | -.(7)638 1298 500 | .(10)502 7613 551 |
| 46.0 | .(4)933 2851 680 | -.(7)604 5338 699 | .(10)466 1016 730 |
| 46.5 | .(4)903 7975 033 | -.(7)573 0383 707 | .(10)432 4654 698 |
| 47.0 | .(4)875 5392 866 | -.(7)543 4917 120 | .(10)401 5750 791 |
| 47.5 | .(4)848 4470 959 | -.(7)515 7546 374 | .(10)373 1808 816 |
| 48.0 | .(4)822 4613 955 | -.(7)489 6991 482 | .(10)347 0582 199 |
| 48.5 | .(4)797 5262 609 | -.(7)465 2074 902 | .(10)323 0046 799 |
| 49.0 | .(4)773 5891 262 | - (7)442 1712 397 | .(10)300 8376 920 |
| 49.5 | .(4)750 6005 505 | -.(7)420 4904 806 | .(10)280 3924 120 |
| 50.0 | (4)728 5140 038 | -.(7)400 0730 601 | .(10)261 5198 458 |

TABLE IV

(The numbers in parentheses denote the number of ciphers between the decimal point and the first significant figure)

| $\rho$ | $A'$ | $B$ | $C$ |
|---|---|---|---|
| 2.0 | 1.000 0000 000 | – 1 250 0000 000 | 250 0000 000 |
| 2.5 | .705 9936 523 | – 495 6054 688 | .(1)615 2343 750 |
| 3.0 | .567 0995 671 | – .265 1515 152 | (1)227 2727 273 |
| 3.5 | .479 4006 348 | – 162 3535 156 | .(1)102 5390 625 |
| 4.0 | .417 2494 172 | – .107 8088 578 | .(2)524 4755 245 |
| 4.5 | .370 3002 930 | –.(1)756 8359 375 | .(2)292 9687 500 |
| 5.0 | .333 3333 333 | –.(1)553 6130 536 | .(2)174 8251 748 |
| 5 5 | .303 3523 560 | –.(1)418 0908 203 | .(2)109 8632 813 |
| 6.0 | .278 4862 197 | –.(1)323 9407 651 | .(3)719 8683 669 |
| 6.5 | .257 4942 453 | –.(1)256 3476 563 | .(3)488 2812 500 |
| 7.0 | .239 5159 021 | – (1)206 4885 579 | .(3)340 9902 791 |
| 7.5 | .223 9329 020 | –.(1)168 8639 323 | .(3)244 1406 250 |
| 8.0 | .210 2881 638 | –.(1)139 9142 653 | .(3)178 6139 557 |
| 8.5 | .198 2357 141 | –.(1)117 2614 820 | (3)133 1676 136 |
| 9.0 | .187 5084 130 | –.(2)992 7311 886 | .(3)100 9557 141 |
| 9.5 | .177 8964 418 | –.(2)848 0187 618 | (4)776 8110 795 |
| 10.0 | 169 2325 443 | –.(2)730 2463 319 | .(4)605 7342 846 |
| 10.5 | .161 3816 481 | – (2)633 3998 033 | .(4)478 0375 874 |
| 11.0 | .154 2334 096 | –.(2)553 0129 672 | .(4)381 3882 532 |
| 11.5 | .147 6967 551 | –.(2)485 7203 344 | .(4)307 3098 776 |
| 12.0 | .141 6958 188 | –.(2)428 9521 906 | (4)249 8750 625 |
| 12.5 | .136 1668 726 | –.(2)380 7227 928 | .(4)204 8732 517 |
| 13 0 | .131 0559 774 | –.(2)339 4807 972 | .(4)169 2702 036 |
| 13.5 | .126 3171 605 | –.(2)304 0020 282 | .(4)140 8503 606 |
| 14 0 | .121 9109 898 | –.(2)273 3115 358 | .(4)117 9762 025 |
| 14 5 | .117 8034 442 | – (2)246 6262 196 | .(5)994 2378 394 |
| 15.0 | .113 9650 116 | –.(2)223 3120 976 | .(5)842 6871 608 |
| 15.5 | 110 3699 628 | –.(2)202 8521 369 | .(5)718 0606 618 |
| 16.0 | .106 9957 611 | – (2)184 8217 922 | .(5)614 9338 741 |
| 16.5 | .103 8225 802 | –.(2)168 8702 311 | .(5)529 0973 297 |
| 17 0 | .100 8329 073 | –.(2)154 7057 999 | .(5)457 2585 218 |
| 17.5 | .(1)980 1121 368 | –.(2)142 0846 788 | .(5)396 8229 973 |
| 18.0 | .(1)953 4368 071 | –.(2)130 8019 601 | .(5)345 7320 530 |
| 18.5 | .(1)928 1796 986 | – (2)120 6845 814 | .(5)302 3413 313 |
| 19.0 | .(1)904 2302 916 | –.(2)111 5856 901 | (5)265 3292 500 |
| 19.5 | .(1)881 4892 929 | –.(2)103 3801 211 | .(5)233 6273 923 |
| 20.0 | .(1)859 8672 410 | –.(3)959 6074 542 | .(5)206 3671 945 |
| 20.5 | .(1)839 2833 163 | –.(3)892 3550 616 | .(5)182 8388 288 |
| 21 0 | .(1)819 6643 189 | –.(3)831 2499 865 | (5)162 4592 807 |
| 21.5 | .(1)800 9437 893 | –.(3)775 6048 512 | .(5)144 7474 061 |
| 22.0 | .(1)783 0612 483 | – (3)724 8225 801 | .(5)129 3043 255 |
| 22.5 | .(1)765 9615 377 | –.(3)678 3828 434 | .(5)115 7979 249 |
| 23.0 | .(1)749 5942 464 | –.(3)635 8307 796 | .(5)103 9505 362 |
| 23 5 | .(1)733 9132 094 | –.(3)596 7675 752 | .(6)935 2909 319 |
| 24.0 | .(1)718 8760 691 | –.(3)560 8425 626 | .(6)843 3722 746 |
| 24.5 | .(1)704 4438 900 | –.(3)527 7465 684 | .(6)762 0889 075 |
| 25.0 | .(1)690 5808 198 | –.(3)497 2062 910 | .(6)690 0318 611 |

TABLE IV—(Continued)

| $\rho$ | $D$ | $E$ | $F$ |
|---|---|---|---|
| 2.0 | 2.454 8611 111 | − .538 1944 444 | .121 5277 778 |
| 2.5 | 586 3715 278 | −.(1)824 6527 778 | .(1)121 5277 778 |
| 3.0 | 214 3308 081 | −.(1)211 4898 990 | .(2)220 9595 960 |
| 3.5 | .(1)962 5552 400 | −.(2)706 2815 657 | .(3)552 3989 899 |
| 4.0 | .(1)491 2101 788 | −.(2)279 2346 542 | .(3)169 9689 200 |
| 4.5 | .(1)274 0445 318 | −.(2)124 4415 307 | .(4)607 0318 570 |
| 5.0 | .(1)163 4129 759 | −.(3)607 0318 570 | .(4)242 8127 428 |
| 5.5 | .(1)102 6452 931 | −.(3)317 9329 351 | .(4)106 2305 750 |
| 6.0 | .(2)672 3770 510 | −.(3)176 3963 161 | .(5)499 9085 881 |
| 6.5 | .(2)455 9791 210 | −.(3)102 6597 994 | .(5)249 9542 941 |
| 7.0 | .(2)318 3891 488 | −.(4)622 0667 018 | .(5)131 5548 916 |
| 7.5 | .(2)227 9369 385 | −.(4)390 2012 053 | .(6)723 5519 039 |
| 8.0 | .(2)166 7477 045 | −.(4)252 2095 208 | .(6)413 4582 308 |
| 8.5 | .(2)124 3142 044 | −.(4)167 3566 157 | .(6)244 3162 273 |
| 9.0 | .(3)942 4020 458 | −.(4)113 6601 579 | .(6)148 7142 253 |
| 9.5 | .(3)725 1166 208 | −.(5)788 0526 136 | .(7)929 4639 082 |
| 10.0 | .(3)565 4114 845 | −.(5)556 6161 004 | .(7)594 8569 012 |
| 10.5 | .(3)446 2073 095 | −.(5)399 7797 905 | .(7)388 9448 970 |
| 11.0 | .(3)355 9882 993 | −.(5)291 5234 609 | .(7)259 2965 980 |
| 11.5 | .(3)286 8401 475 | −.(5)215 5402 971 | .(7)175 9512 629 |
| 12.0 | .(3)233 2288 596 | −.(5)161 3897 791 | .(7)121 3456 986 |
| 12.5 | .(3)191 2235 292 | −.(5)122 2557 913 | .(8)849 4198 899 |
| 13.0 | .(3)157 9915 515 | −.(6)936 0842 049 | .(8)602 8141 154 |
| 13.5 | .(3)131 4646 693 | −.(6)723 8747 984 | .(8)433 2726 455 |
| 14.0 | .(3)110 1142 734 | −.(6)564 9425 143 | .(8)315 1073 785 |
| 14.5 | .(4)927 9784 865 | −.(6)444 6919 780 | .(8)231 6966 018 |
| 15.0 | .(4)786 5252 285 | −.(6)352 8408 251 | .(8)172 1174 757 |
| 15.5 | .(4)670 2028 598 | −.(6)282 0575 132 | .(8)129 0881 067 |
| 16.0 | .(4)573 9480 513 | −.(6)227 0555 132 | .(9)976 8829 690 |
| 16.5 | .(4)493 8316 295 | −.(6)183 9826 864 | .(9)745 5159 507 |
| 17.0 | .(4)426 7803 550 | −.(6)150 0043 633 | .(9)573 4738 083 |
| 17.5 | .(4)370 3725 753 | −.(6)123 0152 522 | .(9)444 4422 014 |
| 18.0 | .(4)322 6867 183 | −.(6)101 4381 253 | .(9)346 8817 182 |
| 18.5 | .(4)282 1879 680 | −.(7)840 8165 075 | .(9)272 5499 214 |
| 19.0 | .(4)247 6427 884 | −.(7)700 3899 143 | .(9)215 5045 890 |
| 19.5 | .(4)218 0539 471 | −.(7)586 1479 930 | .(9)171 4241 049 |
| 20.0 | .(4)192 6107 529 | −.(7)492 7218 558 | .(9)137 1392 839 |
| 20.5 | .(4)170 6506 766 | −.(7)415 9387 633 | .(9)110 3076 849 |
| 21.0 | .(4)151 6295 539 | −.(7)352 5353 142 | .(10)891 8493 673 |
| 21.5 | .(4)135 0983 037 | −.(7)299 9440 718 | .(10)724 6276 109 |
| 22.0 | .(4)120 6846 287 | −.(7)256 1336 780 | .(10)591 5327 436 |
| 22.5 | .(4)108 0785 483 | −.(7)219 4882 245 | .(10)485 0568 497 |
| 23.0 | .(5)970 2089 990 | −.(7)188 7156 473 | .(10)399 4585 821 |
| 23.5 | .(5)872 9415 063 | −.(7)162 7777 261 | .(10)330 3215 199 |
| 24.0 | .(5)787 1501 985 | −.(7)140 8362 750 | .(10)274 2291 863 |
| 24.5 | .(5)711 2852 655 | −.(7)122 2115 427 | .(10)228 5243 219 |
| 25.0 | .(5)644 0316 483 | −.(7)106 3498 773 | .(10)191 1294 329 |

TABLE V

(The numbers in parentheses denote the number of ciphers between the decimal point and the first significant figure)

| $p$ | $A'$ | $B'$ | $C'$ |
|---|---|---|---|
| 2 5 | 2.755 1030 816 | - 1.695 6163 194 | .200 8159 722 |
| 3 0 | 1 170 5555 556 | - 456 9444 444 | .(1)363 8888 889 |
| 3.5 | .658 2671 327 | - .181 6553 455 | (1)104 8944 979 |
| 4.0 | .418 6208 236 | -.(1)868 9782 440 | .(2)382 4786 325 |
| 4 5 | .286 4246 324 | -.(1)466 0481 012 | .(2)162 0459 402 |
| 5.0 | .206 0275 835 | -.(1)270 7119 270 | .(3)763 8888 889 |
| 5.5 | .153 7850 411 | -.(1)166 9293 642 | .(3)390 4384 270 |
| 6.0 | .118 1431 967 | -.(1)107 8702 752 | .(3)212 7325 289 |
| 6.5 | .(1)928 9213 334 | -.(2)724 0120 171 | .(3)122 1004 174 |
| 7.0 | .(1)744 5226 377 | -.(2)501 4851 420 | .(4)731 8541 452 |
| 7.5 | .(1)606 4555 217 | -.(2)356 7180 657 | (4)455 0831 382 |
| 8 0 | .(1)500 8805 006 | -.(2)259 6021 548 | .(4)292 0668 942 |
| 8.5 | .(1)418 6842 624 | -.(2)192 7113 758 | .(4)192 6724 344 |
| 9.0 | .(1)353 6828 101 | -.(2)145 5695 568 | .(4)130 2141 757 |
| 9.5 | .(1)301 5707 729 | -.(2)111 6692 311 | .(5)899 1006 061 |
| 10 0 | .(1)259 2832 573 | -.(3)868 5125 822 | .(5)632 8140 010 |
| 10.5 | .(1)224 5954 598 | -.(3)683 8979 612 | .(5)453 1298 476 |
| 11 0 | .(1)195 8642 756 | -.(3)544 5806 418 | .(5)329 5585 675 |
| 11 5 | .(1)171 8573 720 | -.(3)438 0711 370 | (5)243 1030 464 |
| 12 0 | .(1)151 6376 163 | - (3)355 6755 034 | .(5)181 6613 061 |
| 12.5 | .(1)134 4833 894 | -.(3)291 2423 953 | .(5)137 3671 277 |
| 13.0 | .(1)119 8326 681 | -.(3)240 3543 455 | .(5)105 0128 024 |
| 13 5 | .(1)107 2431 572 | -.(3)199 7956 319 | .(6)810 9219 640 |
| 14.0 | (2)963 6345 096 | -.(3)167 1959 380 | .(6)632 0799 809 |
| 14.5 | .(2)869 1189 432 | - (3)140 7878 615 | .(6)496 9749 385 |
| 15 0 | .(2)786 6089 931 | - (3)119 2394 389 | .(6)393 9212 946 |
| 15.5 | .(2)714 2517 761 | -.(3)101 5368 865 | .(6)314 6050 442 |
| 16 0 | .(2)650 5281 970 | -.(4)869 0145 535 | .(6)253 0429 506 |
| 16.5 | .(2)594 1846 661 | -.(4)747 2977 719 | .(6)204 8829 224 |
| 17 0 | .(2)544 1803 654 | -.(4)645 5059 265 | .(6)166 9275 426 |
| 17.5 | .(2)499 6461 807 | -.(4)559 9303 930 | .(6)136 8055 923 |
| 18.0 | .(2)459 8524 699 | -.(4)487 6318 743 | .(6)112 7430 005 |
| 18.5 | (2)424 1835 784 | - (4)426 2652 106 | (7)934 0140 025 |
| 19.0 | .(2)392 1175 513 | -.(4)373 9474 673 | .(7)777 6318 298 |
| 19.5 | .(2)363 2098 774 | -.(4)329 1578 113 | .(7)650 4887 492 |
| 20.0 | .(2)337 0803 860 | -.(4)290 6609 405 | (7)546 5721 085 |
| 20 5 | .(2)313 4026 161 | -.(4)257 4481 065 | .(7)461 2129 835 |
| 21.0 | (2)291 8951 572 | -.(4)228 6913 803 | .(7)390 7629 285 |
| 21.5 | .(2)272 2750 483 | - (4)203 7079 595 | .(7)332 3537 754 |
| 22.0 | .(2)254 4494 241 | - (4)181 9321 409 | .(7)283 7178 436 |
| 22.5 | .(2)238 1156 895 | -.(4)162 8931 866 | .(7)243 0524 324 |
| 23 0 | .(2)223 1524 850 | -.(4)146 1977 446 | (7)208 9170 140 |
| 23 5 | (2)209 4188 368 | -.(4)131 5158 166 | .(7)180 1547 432 |
| 24 0 | .(2)196 7908 418 | -.(4)118 5694 971 | .(7)155 8321 715 |
| 24 5 | .(2)186 2576 776 | -.(4)107 7593 614 | .(7)135 9946 487 |
| 25.0 | .(2)174 4277 221 | -.(5)969 7978 581 | .(7)117 6202 932 |

TABLE V—(Continued)

| $\rho$ | $D'$ | $E'$ | $C'$ |
|---|---|---|---|
| 2.5 | 1.151 0416 667 | – .140 9722 222 | .(1)175 0000 000 |
| 3.0 | .203 1250 000 | –.(1)170 1388 889 | .(2)145 8333 333 |
| 3.5 | .(1)579 2905 012 | –.(2)355 2350 427 | .(3)224 3589 744 |
| 4.0 | .(1)210 1544 289 | –.(3)988 2478 632 | .(4)480 7692 308 |
| 4 5 | .(2)887 9662 005 | – (3)331 1965 812 | .(4)128 2051 282 |
| 5.0 | .(2)417 9414 336 | –.(3)126 8696 582 | .(5)400 6410 256 |
| 5 5 | .(2)213 4163 324 | –.(4)538 1158 874 | .(5)141 4027 149 |
| 6.0 | .(2)116 2108 929 | –.(4)247 4547 511 | .(6)549 8994 470 |
| 6.5 | (3)666 7389 843 | –.(4)121 5567 199 | .(6)231 5366 092 |
| 7 0 | .(3)399 5246 884 | –.(5)630 9372 602 | .(6)104 1914 742 |
| 7.5 | .(3)248 3850 322 | –.(5)343 1703 316 | .(7)496 1498 770 |
| 8.0 | .(3)159 3881 480 | –.(5)194 3253 685 | .(7)248 0749 385 |
| 8.5 | .(3)105 1352 375 | –.(5)113 9706 601 | .(7)129 4304 027 |
| 9 0 | .(4)710 4822 114 | –.(6)689 3966 587 | .(8)701 0813 479 |
| 9.5 | .(4)490 5434 324 | –.(6)428 5943 973 | .(8)392 6055 548 |
| 10.0 | .(4)345 2433 425 | –.(6)273 0621 968 | .(8)226 5032 047 |
| 10.5 | .(4)247 2044 428 | –.(6)177 8469 607 | .(8)134 2241 213 |
| 11.0 | (4)179 7851 483 | –.(6)118 1651 639 | .(9)814 9321 650 |
| 11.5 | .(4)132 6177 560 | –.(7)799 4765 550 | .(9)505 8199 645 |
| 12.0 | .(5)990 9817 818 | –.(7)549 9387 059 | .(9)320 3526 442 |
| 12 5 | .(5)749 3410 036 | –.(7)384 0787 078 | .(9)206 6791 253 |
| 13.0 | .(5)572 8402 600 | –.(7)272 0198 696 | .(9)135 6331 760 |
| 13 5 | .(5)442 3498 419 | –.(7)195 1610 699 | .(10)904 2211 731 |
| 14.0 | .(5)344 7903 612 | –.(7)141 7056 417 | .(10)611 6790 289 |
| 14 5 | .(5)271 0905 849 | –.(7)104 0436 901 | .(10)419 4370 484 |
| 15 0 | .(5)214 8754 286 | –.(8)771 8806 793 | .(10)291 2757 280 |
| 15.5 | .(5)171 6092 639 | –.(8)578 2216 817 | .(10)204 6802 413 |
| 16 0 | .(5)138 0280 549 | –.(8)437 1000 417 | .(10)145 4306 978 |
| 16.5 | .(5)111 7576 435 | –.(8)333 2476 075 | .(10)104 4117 830 |
| 17 0 | .(6)910 5379 546 | –.(8)256 1134 028 | .(11)756 9854 269 |
| 17.5 | .(6)746 2299 456 | –.(8)198 3240 275 | .(11)553 8917 758 |
| 18 0 | .(6)614 9748 289 | –.(8)154 6720 805 | .(11)408 8248 821 |
| 18.5 | .(6)509 4718 636 | –.(8)121 4431 743 | .(11)304 2417 727 |
| 19 0 | .(6)424 1700 765 | –.(9)959 6292 582 | .(11)228 1813 296 |
| 19.5 | (6)354 8175 306 | –.(9)762 8862 452 | .(11)172 4036 712 |
| 20 0 | .(6)298 1344 148 | –.(9)609 9716 846 | .(11)131 1767 064 |
| 20 5 | .(6)251 5739 152 | –.(9)490 3776 023 | .(11)100 4757 751 |
| 21.0 | .(6)213 1458 799 | –.(9)396 2862 254 | .(12)774 5007 663 |
| 21 5 | .(6)181 2857 886 | –.(9)321 8393 150 | .(12)600 6332 473 |
| 22 0 | .(6)154 7566 781 | –.(9)262 6168 770 | .(12)468 4939 329 |
| 22.5 | .(6)132 5752 232 | –.(9)215 2622 450 | .(12)367 4462 219 |
| 23.0 | .(6)113 9556 568 | –.(9)177 2103 622 | .(12)289 7172 134 |
| 23.5 | .(7)982 6695 408 | –.(9)146 4894 049 | .(12)229 5872 257 |
| 24.0 | .(7)849 9994 800 | –.(9)121 5749 393 | .(12)182 8194 575 |
| 24.5 | .(7)741 7936 742 | –.(9)101 8827 904 | .(12)147 1231 630 |
| 25.0 | .(7)641 5689 752 | –.(10)846 8458 443 | .(12)117 5267 941 |

TABLE VI

(The numbers in parentheses denote the number of ciphers between the decimal point and the first significant figure.)

| $\rho$ | $A$ | $B$ |
|---|---|---|
| 3.0 | 1.000 0000 000 | -1.361 1111 111 |
| 3 5 | .745 3327 179 | - 617 5664 266 |
| 4.0 | 619 2696 193 | -.362 6910 127 |
| 4 5 | .536 5078 449 | - 238 2141 749 |
| 5 0 | .475 9358 289 | -.167 1854 290 |
| 5.5 | .428 9313 952 | -.122 7793 517 |
| 6 0 | .391 0671 372 | -.(1)932 5031 693 |
| 6 5 | .359 7514 629 | - (1)727 0467 546 |
| 7.0 | .333 3333 333 | -.(1)578 9958 809 |
| 7.5 | .310 6966 019 | -.(1)469 2645 603 |
| 8 0 | .291 0527 351 | -.(1)386 0218 617 |
| 8 5 | .273 8253 011 | -.(1)321 6238 375 |
| 9.0 | 258 5812 357 | -.(1)270 9606 497 |
| 9.5 | 244 9877 912 | -.(1)230 5173 392 |
| 10.0 | .232 7844 334 | - (1)197 8161 991 |
| 10 5 | .221 7638 626 | -.(1)171 0732 926 |
| 11.0 | .211 7588 265 | -.(1)148 9801 195 |
| 11.5 | .202 6327 307 | -.(1)130 5610 007 |
| 12 0 | .194 2728 127 | -.(1)115 0776 715 |
| 12 5 | .186 5850 870 | -.(1)101 9640 947 |
| 13.0 | .179 4905 415 | - (2)907 8107 302 |
| 13.5 | .172 9222 326 | - (2)811 8410 815 |
| 14.0 | .166 8230 401 | -.(2)729 0028 821 |
| 14.5 | 161 1439 098 | -.(2)657 1143 441 |
| 15.0 | .155 8424 640 | -.(2)594 4165 421 |
| 15.5 | .150 8818 928 | -.(2)539 4804 055 |
| 16.0 | 146 2300 606 | -.(2)491 1364 663 |
| 16.5 | .141 8587 811 | -.(2)448 4212 624 |
| 17.0 | .137 7432 239 | -.(2)410 5360 674 |
| 17.5 | 133 8614 260 | -.(2)376 8148 358 |
| 18 0 | .130 1938 866 | -.(2)346 6991 057 |
| 18.5 | 126 7232 293 | -.(2)319 7181 992 |
| 19.0 | .123 4339 187 | -.(2)295 4734 917 |
| 19.5 | .120 3120 213 | -.(2)273 6258 312 |
| 20.0 | .117 3450 034 | -.(2)253 8854 120 |
| 20 5 | .114 5607 471 | -.(2)236 0035 778 |
| 21.0 | .111 8314 598 | -.(2)219 7661 473 |
| 21.5 | .109 2654 341 | -.(2)204 9879 526 |
| 22.0 | .106 8150 524 | -.(2)191 5083 451 |
| 22.5 | .104 4726 355 | -.(2)179 1874 832 |
| 23.0 | .102 2311 723 | -.(2)167 9032 479 |
| 23.5 | .100 0842 483 | - (2)157 5486 730 |
| 24.0 | .(1)980 2598 324 | -.(2)148 0297 913 |
| 24.5 | .(1)960 5097 234 | -.(2)139 2638 202 |
| 25.0 | .(1)941 5425 829 | -.(2)131 1776 633 |

TABLE VI—(Continued)

| $\rho$ | $C$ | $D$ |
|---|---|---|
| 3.0 | .388 8888 889 | – (1)277 7777 778 |
| 3.5 | .116 7093 913 | –.(2)581 8684 896 |
| 4.0 | .(1)498 5754 986 | –.(2)185 1851 852 |
| 4.5 | .(1)251 6301 473 | –.(3)727 3356 120 |
| 5.0 | .(1)140 7742 584 | –.(3)326 7973 856 |
| 5.5 | .(2)846 3824 237 | –.(3)161 6301 360 |
| 6.0 | .(2)537 1649 335 | –.(4)859 9931 201 |
| 6.5 | .(2)355 7417 128 | – (4)484 8904 080 |
| 7.0 | .(2)243 8852 284 | –.(4)286 6643 734 |
| 7 5 | .(2)172 0852 322 | –.(4)176 3237 847 |
| 8 0 | .(2)124 4257 604 | – (4)112 1730 157 |
| 8.5 | (3)918 7769 007 | –.(5)734 6824 363 |
| 9 0 | .(3)690 9857 765 | –.(5)493 5612 689 |
| 9.5 | .(3)528 1236 437 | –.(5)339 0842 014 |
| 10.0 | .(3)409 4730 527 | –.(5)237 6406 110 |
| 10.5 | .(3)321 5757 191 | –.(5)169 6421 007 |
| 11.0 | .(3)255 4794 154 | –.(5)122 9175 574 |
| 11 5 | .(3)205 1024 695 | –.(6)904 2245 370 |
| 12.0 | .(3)166 2345 586 | –.(6)674 0640 244 |
| 12.5 | .(3)135 9108 167 | –.(6)508 6263 021 |
| 13.0 | .(3)112 0109 002 | – (6)388 0974 686 |
| 13.5 | .(4)929 9691 087 | –.(6)299 1919 424 |
| 14 0 | .(4)777 3890 832 | –.(6)232 8584 812 |
| 14.5 | .(4)653 9677 383 | –.(6)182 8395 204 |
| 15.0 | .(4)513 3898 749 | –.(6)144 7498 667 |
| 15.5 | .(4)470 8598 806 | –.(6)115 4775 918 |
| 16.0 | .(4)402 7015 521 | –.(7)927 8837 607 |
| 16 5 | .(4)346 0719 079 | –.(7)750 6043 467 |
| 17.0 | .(4)298 7541 988 | –.(7)611 0454 032 |
| 17.5 | .(4)259 0066 153 | –.(7)500 4028 978 |
| 18.0 | .(4)225 4506 124 | –.(7)412 1003 883 |
| 18.5 | .(4)196 9877 124 | –.(7)341 1837 940 |
| 19.0 | .(4)172 7369 850 | –.(7)283 8913 786 |
| 19.5 | .(4)151 9876 810 | –.(7)237 3452 480 |
| 20.0 | .(4)134 1630 697 | –.(7)199 3279 892 |
| 20.5 | .(4)118 7926 279 | –.(7)168 1195 507 |
| 21.0 | .(4)105 4905 051 | –.(7)142 3771 352 |
| 21.5 | .(5)939 3873 877 | –.(7)121 0460 765 |
| 22.0 | .(5)838 7409 116 | –.(7)103 2932 157 |
| 22.5 | .(5)750 7766 502 | –.(8)884 5674 819 |
| 23.0 | .(5)673 6666 593 | –.(8)760 0821 534 |
| 23 5 | .(5)605 8783 224 | –.(8)655 2351 718 |
| 24.0 | .(5)546 1216 848 | –.(8)566 6066 961 |
| 24.5 | .(5)493 3070 093 | –.(8)491 4263 585 |
| 25.0 | .(5)446 5105 454 | –.(8)427 4401 392 |

TABLE VI—(Continued)

| $\rho$ | $E$ | $F$ |
|---|---|---|
| 3.0 | 2.988 9351 852 | −.948 1481 481 |
| 3.5 | .875 4725 025 | − 189 2894 604 |
| 4.0 | .370 6973 366 | −.(1)590 7882 241 |
| 4.5 | .186 3394 083 | −.(1)229 8587 984 |
| 5.0 | .104 0300 478 | −.(1)102 7515 921 |
| 5.5 | .(1)624 7216 370 | −.(2)506 6826 853 |
| 6 0 | .(1)396 1994 936 | −.(2)269 0942 361 |
| 6.5 | .(1)262 2649 076 | −.(2)151 5410 433 |
| 7.0 | .(1)179 7450 169 | −.(3)895 1744 340 |
| 7.5 | .(1)126 8009 428 | −.(3)550 2997 477 |
| 8.0 | .(2)916 6922 412 | −.(3)349 9462 528 |
| 8.5 | .(2)676 8238 027 | −.(3)229 1312 716 |
| 9 0 | .(2)508 9783 157 | −.(3)153 8970 152 |
| 9 5 | .(2)388 9905 578 | −.(3)105 7119 802 |
| 10.0 | .(2)301 5839 811 | −.(4)740 7668 221 |
| 10.5 | .(2)236 8373 742 | −.(4)528 4390 327 |
| 11.0 | .(2)188 1526 456 | − (4)383 0865 153 |
| 11.5 | .(2)151 0481 065 | −.(4)281 7940 194 |
| 12.0 | .(2)122 4214 851 | −.(4)210 0557 062 |
| 12 5 | .(2)100 0884 111 | −.(4)158 4944 724 |
| 13.0 | .(3)824 6707 418 | −.(4)120 9320 228 |
| 13.5 | .(3)684 8389 962 | − (5)932 2623 816 |
| 14.0 | .(3)572 4725 971 | −.(5)725 5546 509 |
| 14.5 | .(3)481 5811 670 | −.(5)569 6913 075 |
| 15 0 | .(3)407 5132 846 | −.(5)451 0040 878 |
| 15.5 | .(3)346 7368 814 | −.(5)359 7939 891 |
| 16.0 | .(3)296 5444 286 | −.(5)289 0976 344 |
| 16.5 | .(3)254 8421 133 | −.(5)233 8608 712 |
| 17.0 | .(3)219 9973 645 | −.(5)190 3777 170 |
| 17.5 | .(3)190 7274 117 | −.(5)155 9046 664 |
| 18.0 | .(3)166 0170 278 | −.(5)128 3924 319 |
| 18.5 | .(3)145 0572 508 | − (5)106 2973 088 |
| 19.0 | .(3)127 1993 361 | −.(6)884 4715 492 |
| 19.5 | .(3)111 9198 707 | −.(6)739 4524 672 |
| 20.0 | .(4)987 9413 925 | − (6)621 0067 139 |
| 20.5 | .(4)874 7564 250 | −.(6)523 7749 416 |
| 21.0 | .(4)776 8023 679 | −.(6)443 5733 082 |
| 21.5 | .(4)691 6366 980 | −.(6)377 1157 568 |
| 22.0 | (4)617 6240 310 | −.(6)321 8064 139 |
| 22.5 | .(4)552 8493 126 | −.(6)275 5833 129 |
| 23.0 | .(4)496 0674 883 | −.(6)236 7999 628 |
| 23 5 | .(4)446 1499 515 | −.(6)204 1350 268 |
| 24.0 | .(4)402 1467 932 | −.(6)176 5230 172 |
| 24.5 | .(4)365 3958 471 | −.(6)154 0028 718 |
| 25.0 | .(4)328 7959 685 | −.(6)133 1661 235 |

TABLE VI—(Continued)

| ρ | G | H |
|---|---|---|
| 3.0 | .(1)703 2407 407 | 311 9212 963 |
| 3.5 | .(2)996 0214 120 | .(1)431 6767 940 |
| 4.0 | (2)233 6419 753 | .(1)100 2196 106 |
| 4.5 | .(3)711 2449 363 | .(2)303 4820 991 |
| 5.0 | .(3)256 2636 166 | (2)109 0241 118 |
| 5.5 | (3)104 2151 284 | .(3)442 3904 068 |
| 6.0 | .(4)464 8740 588 | .(3)197 2080 911 |
| 6.5 | .(4)223 1993 819 | (4)916 1621 940 |
| 7.0 | .(4)113 8216 820 | .(4)482 2607 431 |
| 7.5 | .(5)610 4831 370 | (4)258 5707 479 |
| 8.0 | .(5)341 8161 061 | .(4)144 7403 415 |
| 8.5 | .(5)198 6271 380 | .(5)840 9223 280 |
| 9.0 | .(5)119 2274 520 | .(5)504 7006 725 |
| 9.5 | .(6)736 4557 705 | .(5)311 7158 600 |
| 10.0 | .(6)466 6351 290 | .(5)197 4943 272 |
| 10.5 | .(6)302 4952 338 | .(5)128 0172 024 |
| 11.0 | .(6)200 1684 248 | .(6)817 0785 482 |
| 11.5 | .(6)134 9506 134 | .(6)571 0651 132 |
| 12.0 | .(7)925 4160 570 | (6)391 5919 775 |
| 12.5 | .(7)644 5499 552 | .(6)272 7356 553 |
| 13.0 | (7)455 3925 745 | (6)192 6912 065 |
| 13.5 | .(7)326 0189 201 | .(6)137 9465 952 |
| 14.0 | .(7)236 2653 096 | (7)999 6815 163 |
| 14.5 | .(7)173 1717 709 | .(7)732 7121 006 |
| 15.0 | .(7)128 2726 524 | .(7)542 7324 021 |
| 15.5 | .(8)959 5450 485 | (7)405 9879 530 |
| 16.0 | (8)724 4284 089 | .(7)306 5060 664 |
| 16.5 | .(8)551 6645 046 | .(7)233 4084 850 |
| 17.0 | .(8)423 5236 197 | (7)179 1912 403 |
| 17.5 | .(8)327 6400 420 | .(7)138 6225 823 |
| 18.0 | .(8)255 2963 323 | (7)108 0139 227 |
| 18.5 | .(8)200 2846 930 | (8)847 3860 468 |
| 19.0 | .(8)158 1422 442 | (8)669 0830 603 |
| 19.5 | (8)125 6316 099 | .(8)531 5324 962 |
| 20.0 | (8)100 3843 545 | (8)424 7132 721 |
| 20.5 | .(9)806 5374 215 | .(8)341 2348 327 |
| 21.0 | (9)651 4166 128 | .(8)275 6048 185 |
| 21.5 | .(9)528 7636 968 | .(8)223 7117 748 |
| 22.0 | (9)431 2539 387 | .(8)182 4566 123 |
| 22.5 | .(9)353 3296 685 | .(8)149 4878 961 |
| 23.0 | .(9)290 7474 080 | .(8)123 0102 298 |
| 23.5 | .(9)240 2476 147 | .(8)101 6445 207 |
| 24.0 | .(9)199 3121 346 | .(9)843 2535 668 |
| 24.5 | .(9)166 9628 406 | (9)706 3889 111 |
| 25.0 | .(9)138 7382 906 | .(9)586 9755 711 |

TABLE VI—(Continued)

| $\rho$ | $I$ | $J$ |
|---|---|---|
| 3.0 | -.(1)234 9537 037 | .(2)178 2407 407 |
| 3.5 | -.(2)232 9282 407 | .(3)127 3148 148 |
| 4.0 | -.(3)408 9506 173 | .(4)169 7530 864 |
| 4 5 | -.(4)972 9456 019 | .(5)318 2870 370 |
| 5.0 | -.(4)282 5435 730 | (6)748 9106 754 |
| 5 5 | -.(5)947 9582 728 | (6)208 0307 432 |
| 6.0 | -.(5)355 3443 795 | .(7)656 9391 889 |
| 6.5 | -.(5)145 5344 260 | .(7)229 9287 161 |
| 7.0 | - (6)641 0133 904 | .(8)875 9189 186 |
| 7 5 | -.(6)300 1017 659 | (8)358 3304 667 |
| 8.0 | -.(6)148 0060 623 | (8)155 7958 551 |
| 8 5 | - (7)763 5619 773 | .(9)714 0643 358 |
| 9.0 | -.(7)409 7430 989 | .(9)342 7508 812 |
| 9.5 | -.(7)227 6566 932 | .(9)171 3754 406 |
| 10 0 | -.(7)130 4646 031 | .(10)888 6133 957 |
| 10.5 | -.(8)768 7010 766 | (10)476 0428 905 |
| 11 0 | -.(8)464 4029 703 | .(10)262 6443 534 |
| 11.5 | -.(8)287 0085 966 | .(10)148 8318 003 |
| 12 0 | -.(8)181 0859 646 | .(11)864 1846 467 |
| 12.5 | -.(8)116 4409 022 | .(11)513 1096 340 |
| 13 0 | -.(9)761 8900 625 | .(11)310 9755 357 |
| 13.5 | -.(9)506 5928 672 | .(11)192 0731 250 |
| 14 0 | -.(9)341 8901 625 | .(11)120 7316 786 |
| 14.5 | -.(9)233 9443 041 | .(12)771 3412 798 |
| 15.0 | -.(9)162 1522 356 | (12)500 3294 788 |
| 15.5 | -.(9)113 7486 502 | .(12)329 1641 308 |
| 16.0 | -(10)806 9508 539 | .(12)219 4427 539 |
| 16 5 | -(10)578 5246 623 | (12)148 1238 589 |
| 17 0 | -(10)418 8850 767 | (12)101 1577 573 |
| 17 5 | -(10)306 1363 264 | .(13)698 4702 287 |
| 18.0 | -(10)225 7107 282 | .(13)487 3048 107 |
| 18.5 | -(10)167 8017 503 | (13)343 3283 894 |
| 19.0 | -(10)125 7344 857 | .(13)244 1446 325 |
| 19.5 | -(11)949 1786 023 | .(13)175 1472 363 |
| 20 0 | -(11)721 6269 402 | .(13)126 7022 561 |
| 20 5 | -(11)552 3276 496 | .(14)923 8706 171 |
| 21.0 | -(11)425 4604 167 | .(14)678 7620 861 |
| 21.5 | -(11)329 7379 935 | (14)502 2839 437 |
| 22.0 | -(11)257 0422 413 | .(14)374 2507 816 |
| 22.5 | -(11)201 4893 910 | .(14)280 6880 862 |
| 23.0 | -(11)158 7837 578 | .(14)211 8400 650 |
| 23.5 | -(11)125 7671 107 | .(14)160 8415 309 |
| 24 0 | -(11)100 1019 200 | .(14)122 8244 418 |
| 24 5 | -(12)805 1862 542 | (15)948 6730 535 |
| 25 0 | -(12)642 9736 393 | .(15)728 0195 608 |

TABLE VII

(The numbers in parentheses denote the number of ciphers between the decimal point and the first significant figure)

| $\rho$ | $A'$ | $B'$ |
|---|---|---|
| 3 5 | 28.751 2015 275 | -2.018 4913 853 |
| 4.0 | 1.362 9280 045 | -.639 9553 571 |
| 4.5 | .826 9536 776 | - 286 6103 525 |
| 5.0 | .557 2123 756 | -.150 4344 206 |
| 5.5 | .399 2748 295 | -.(1)869 6309 477 |
| 6 0 | 298 3433 421 | -.(1)537 5591 718 |
| 6.5 | 229 9463 348 | -.(1)349 3835 704 |
| 7.0 | .181 5696 107 | - (1)236 2100 221 |
| 7 5 | .146 2040 440 | - (1)164 8961 618 |
| 8.0 | 119 6560 886 | - (1)118 2279 748 |
| 8.5 | (1)992 8681 536 | -.(2)867 1239 512 |
| 9 0 | .(1)833 6752 283 | - (2)648 5413 101 |
| 9 5 | .(1)707 2793 855 | -.(2)493 4138 461 |
| 10.0 | .(1)605 5349 149 | -.(2)381 0899 486 |
| 10 5 | (1)522 6404 211 | -.(2)298 3079 274 |
| 11 0 | .(1)454 3790 237 | -.(2)236 3306 957 |
| 11.5 | .(1)397 6190 688 | -.(2)189 2706 474 |
| 12 0 | .(1)350 0205 660 | -.(2)153 0798 651 |
| 12.5 | (1)309 7896 316 | -.(2)124 9247 269 |
| 13.0 | .(1)275 5429 712 | -.(2)102 7891 576 |
| 13.5 | .(1)246 1999 134 | -.(3)852 1744 589 |
| 14 0 | .(1)220 9073 683 | -.(3)711 4425 518 |
| 14.5 | (1)198 9854 583 | -.(3)597 8026 293 |
| 15 0 | .(1)179 8875 910 | - (3)505 3397 258 |
| 15.5 | .(1)163 1707 725 | -.(3)429 5745 974 |
| 16 0 | .(1)148 4732 835 | -.(3)367 0820 942 |
| 16.5 | (1)135 4977 156 | - (3)315 2193 163 |
| 17 0 | .(1)123 9979 553 | -.(3)271 9296 956 |
| 17.5 | .(1)113 7691 065 | -.(3)235 6002 537 |
| 18 0 | .(1)104 6396 241 | -.(3)204 9565 434 |
| 18.5 | .(2)964 6512 317 | -.(3)178 9845 869 |
| 19 0 | (2)891 2347 318 | -.(3)156 8723 569 |
| 19.5 | .(2)842 5297 627 | -.(3)137 9655 374 |
| 20 0 | .(2)765 3875 595 | -.(3)121 7338 175 |
| 20 5 | .(2)711 3114 773 | -.(3)107 7450 175 |
| 21.0 | .(2)662 2270 210 | -.(4)956 4508 777 |
| 21 5 | (2)617 5695 829 | - (4)851 4293 652 |
| 22.0 | .(2)576 8498 581 | -.(4)759 9627 856 |
| 22 5 | .(2)539 6419 780 | - (4)680 0595 916 |
| 23.0 | .(2)505 5744 586 | - (4)610 0441 976 |
| 23 5 | .(2)474 3260 497 | -.(4)548 5160 065 |
| 24 0 | .(2)445 5992 272 | -.(4)494 2969 537 |
| 24 5 | .(2)419 1587 794 | - (4)446 3969 384 |
| 25.0 | .(2)394 7661 411 | -.(4)403 9598 359 |

TABLE VII—(Continued)

| $\rho$ | $C'$ | $D'$ |
|---|---|---|
| 3.5 | 357 6197 452 | – (1)173 0659 191 |
| 4.0 | (1)798 5119 048 | – (2)282 3837 868 |
| 4.5 | .(1)269 7269 968 | –.(3)732 8051 663 |
| 5.0 | .(1)111 5137 722 | –.(3)241 1381 219 |
| 5.5 | .(2)523 4528 750 | –.(4)925 0558 091 |
| 6.0 | .(2)268 7939 211 | –.(4)396 2769 318 |
| 6.5 | .(2)147 7366 521 | –.(4)184 7282 763 |
| 7.0 | (3)856 9155 732 | – (5)921 2018 141 |
| 7.5 | .(3)519 3903 212 | –.(5)485 5539 780 |
| 8 0 | .(3)326 6117 660 | –.(5)268 1183 076 |
| 8.5 | .(3)211 9254 433 | – (5)154 0554 574 |
| 9.0 | .(3)141 2859 194 | –.(6)916 1750 131 |
| 9 5 | .(4)964 5009 167 | –.(6)561 5350 601 |
| 10.0 | .(4)672 3458 517 | –.(6)353 4749 049 |
| 10.5 | .(4)477 5025 292 | –.(6)227 8602 485 |
| 11.0 | .(4)344 8411 857 | – (6)150 0558 578 |
| 11.5 | (4)252 8236 420 | –.(6)100 7434 807 |
| 12.0 | (4)187 9176 343 | –.(7)688 3248 748 |
| 12.5 | (4)141 4318 835 | –.(7)477 8804 517 |
| 13 0 | .(4)107 6721 454 | –.(7)336 6794 370 |
| 13.5 | .(5)828 3962 809 | –.(7)240 4245 028 |
| 14 0 | .(5)643 5779 618 | –.(7)173 8435 375 |
| 14.5 | .(5)504 5238 369 | –.(7)127 1625 885 |
| 15 0 | .(5)398 8437 502 | –.(8)940 2155 070 |
| 15 5 | .(5)317 7726 592 | –.(8)702 1757 322 |
| 16.0 | .(5)255 0351 049 | –.(8)529 3336 632 |
| 16.5 | .(5)206 0875 950 | – (8)402 5511 467 |
| 17.0 | .(5)167 6061 262 | – (8)308 6648 353 |
| 17 5 | (5)137 1349 784 | –.(8)238 5149 470 |
| 18.0 | .(5)112 8433 182 | –.(8)185 6576 541 |
| 18.5 | .(6)933 5432 357 | –.(8)145 5130 248 |
| 19.0 | .(6)776 2421 574 | – (8)114 7942 689 |
| 19 5 | (6)648 5562 423 | –.(9)911 2100 394 |
| 20.0 | .(6)544 3500 271 | – (9)727 5436 127 |
| 20 5 | .(6)458 8701 286 | –.(9)584 1368 200 |
| 21.0 | .(6)388 4099 727 | – (9)471 4844 101 |
| 21.5 | .(6)330 0128 092 | –.(9)382 4793 533 |
| 22.0 | .(6)281 5293 946 | –.(9)311 7704 559 |
| 22.5 | .(6)240 9925 109 | –.(9)255 3016 408 |
| 23.0 | .(6)206 9976 250 | –.(9)209 9789 105 |
| 23.5 | (6)178 3794 880 | – (9)173 4277 959 |
| 24 0 | .(6)154 1991 740 | –.(9)143 8154 281 |
| 24 5 | .(6)133 6979 531 | – (9)119 7203 889 |
| 25.0 | .(6)116 2537 745 | –.(9)100 0289 166 |

TABLE VII—(Continued)

| $\rho$ | $E'$ | $F'$ |
|---|---|---|
| 3.5 | 1.579 8913 122 | – 291 1769 387 |
| 4.0 | .344 9276 620 | – (1)455 1504 630 |
| 4 5 | .115 4442 217 | –.(1)115 8781 297 |
| 5.0 | .(1)475 1410 972 | –.(2)377 5757 988 |
| 5.5 | .(1)222 4862 588 | –.(2)144 0406 919 |
| 6.0 | .(1)114 0808 001 | –.(3)614 9607 748 |
| 6.5 | .(2)626 4468 334 | –.(3)286 0508 258 |
| 7 0 | .(2)363 1390 089 | –.(3)142 4423 327 |
| 7.5 | (2)220 0141 501 | –.(4)750 0507 440 |
| 8.0 | .(2)138 3131 064 | –.(4)413 8794 788 |
| 8.5 | .(3)897 2722 669 | – (4)237 6853 941 |
| 9.0 | .(3)598 0994 308 | – (4)141 2990 508 |
| 9.5 | .(3)408 2508 260 | –.(5)865 7917 011 |
| 10 0 | .(3)284 5633 018 | –.(5)544 8786 925 |
| 10 5 | .(3)202 0841 107 | –.(5)351 1847 710 |
| 11.0 | (3)145 9325 974 | –.(5)231 2393 114 |
| 11.5 | .(3)106 9873 276 | – (5)155 2312 455 |
| 12.0 | .(4)795 1834 088 | –.(5)106 0518 396 |
| 12.5 | .(4)598 4598 613 | –.(6)736 2301 182 |
| 13.0 | .(4)455 5973 535 | –.(6)518 6638 399 |
| 13.5 | .(4)350 5159 780 | –.(6)370 3629 282 |
| 14.0 | .(4)272 3103 032 | –.(6)267 7874 363 |
| 14.5 | .(4)213 4710 252 | –.(6)195 8739 375 |
| 15.0 | (4)168 7544 633 | –.(6)144 8213 551 |
| 15.5 | .(4)134 4512 878 | –.(6)108 1535 449 |
| 16.0 | .(4)107 9058 431 | –.(7)815 2967 606 |
| 16.5 | .(5)871 9546 641 | –.(7)620 0115 559 |
| 17.0 | .(5)709 1357 523 | –.(7)475 4002 735 |
| 17.5 | .(5)580 2104 892 | –.(7)367 3519 091 |
| 18.0 | .(5)477 4317 347 | –.(7)285 9398 751 |
| 18.5 | .(5)394 9737 353 | – (7)224 1091 135 |
| 19.0 | .(5)328 4199 721 | –.(7)176 7967 053 |
| 19 5 | .(5)274 3965 893 | –.(7)140 3360 406 |
| 20.0 | (5)230 3075 611 | –.(7)112 0487 209 |
| 20.5 | .(5)194 1416 638 | –.(8)899 6217 202 |
| 21.0 | (5)164 3306 063 | – (8)726 1234 346 |
| 21 5 | .(5)139 6233 668 | –.(8)589 0458 851 |
| 22.0 | .(5)119 1105 861 | –.(8)480 1471 084 |
| 22 5 | (5)101 9599 333 | – (8)393 1799 685 |
| 23.0 | .(6)875 7714 928 | –.(8)323 3791 671 |
| 23.5 | .(6)754 6921 922 | –.(8)267 0876 406 |
| 24.0 | .(6)652 3888 702 | –.(8)221 4825 166 |
| 24.5 | .(6)565 6513 791 | – (8)184 3745 724 |
| 25 0 | .(6)491 8478 631 | –.(8)154 0485 226 |

TABLE VII—(Continued)

| $\rho$ | $G'$ | $H'$ |
|---|---|---|
| 3.5 | .(1)143 3986 442 | .(1)545 8043 981 |
| 4 0 | .(2)165 3852 513 | .(2)616 8981 481 |
| 4 5 | .(3)325 3719 022 | .(2)120 1878 234 |
| 5 0 | .(4)847 0633 624 | .(3)311 2518 155 |
| 5.5 | .(4)264 7923 510 | .(4)969 9931 081 |
| 6.0 | .(5)944 9259 723 | .(4)345 4902 916 |
| 6.5 | .(5)373 3302 227 | .(4)136 3314 118 |
| 7.0 | .(5)160 0116 167 | .(5)583 8397 737 |
| 7.5 | .(6)733 3626 200 | .(5)267 4300 767 |
| 8.0 | .(6)355 6040 392 | .(5)129 6221 514 |
| 8.5 | .(6)180 9471 484 | .(6)659 3767 449 |
| 9 0 | .(7)960 0363 156 | .(6)349 7616 94[illegible] |
| 9 5 | .(7)528 3674 230 | .(6)192 4628 42[illegible] |
| 10.0 | .(7)300 3768 815 | .(6)109 4008 386 |
| 10.5 | .(7)175 7761 725 | .(7)640 1325 618 |
| 11.0 | (7)105 5698 126 | .(7)384 4277 647 |
| 11.5 | .(8)649 1137 736 | .(7)236 3567 152 |
| 12.0 | .(8)407 7298 452 | .(7)148 4557 926 |
| 12.5 | .(8)261 1497 436 | .(8)950 8148 611 |
| 13.0 | .(8)170 2826 250 | .(8)619 9571 695 |
| 13.5 | .(8)112 8752 998 | .(8)410 9394 710 |
| 14.0 | .(9)759 6818 749 | .(8)276 5670 073 |
| 14.5 | .(9)518 5449 520 | (8)188 7758 141 |
| 15.0 | .(9)358 6183 357 | .(8)130 5524 791 |
| 15.5 | .(9)251 0639 560 | .(9)913 9675 989 |
| 16 0 | .(9)177 7849 669 | .(9)647 1965 291 |
| 16.5 | .(9)127 2480 223 | .(9)463 2203 741 |
| 17 0 | .(10)919 9584 852 | .(9)334 8890 426 |
| 17 5 | .(10)671 4105 266 | .(9)244 4091 600 |
| 18 0 | .(10)494 3960 450 | .(9)179 9705 267 |
| 18.5 | .(10)367 1241 847 | .(9)133 6401 141 |
| 19.0 | .(10)274 7920 816 | .(9)100 0289 750 |
| 19.5 | .(10)207 2366 516 | .(10)754 3731 214 |
| 20 0 | .(10)157 4099 489 | .(10)572 9940 142 |
| 20.5 | .(10)120 3776 663 | .(10)438 1898 319 |
| 21.0 | .(11)926 5399 250 | .(10)337 2711 214 |
| 21.5 | .(11)717 5522 633 | .(10)261 1964 982 |
| 22 0 | .(11)558 9721 931 | .(10)203 4711 966 |
| 22 5 | .(11)437 8836 453 | .(10)159 3934 600 |
| 23 0 | .(11)344 8668 374 | .(10)125 5342 827 |
| 23.5 | (11)273 0031 904 | .(11)993 7516 417 |
| 24.0 | .(11)217 1767 993 | .(11)790 5382 130 |
| 24.5 | .(11)173 5819 878 | .(11)631 8492 407 |
| 25.0 | .(11)139 3623 589 | .(11)507 2869 968 |

TABLE VII—(Continued)

| $\rho$ | $I'$ | $J'$ |
|---|---|---|
| 3.5 | –.(2)270 9986 772 | .(3)135 1095 994 |
| 4.0 | –.(3)227 3478 836 | .(5)844 4349 962 |
| 4.5 | –.(4)343 6965 064 | .(6)993 4529 367 |
| 5.0 | –.(5)713 2482 623 | .(6)165 5754 895 |
| 5 5 | –.(5)182 5352 461 | .(7)348 5799 778 |
| 6.0 | –.(6)544 3210 423 | .(8)871 4499 445 |
| 6.5 | –.(6)182 6693 153 | .(8)248 9856 984 |
| 7 0 | –.(7)674 0025 445 | .(9)792 2272 223 |
| 7.5 | –.(7)268 9333 213 | .(9)275 5572 947 |
| 8 0 | –.(7)114 6212 362 | .(9)103 3339 855 |
| 8.5 | –.(8)516 9083 906 | (10)413 3359 421 |
| 9.0 | –.(8)244 8220 580 | .(10)174 8728 986 |
| 9.5 | –.(8)121 0509 065 | .(11)777 2128 825 |
| 10.0 | –.(9)621 7703 060 | .(11)360 8488 383 |
| 10.5 | –.(9)330 4159 767 | .(11)174 2028 875 |
| 11.0 | –.(9)181 0370 007 | .(12)871 0144 373 |
| 11.5 | –.(9)101 9713 676 | .(12)449 5558 386 |
| 12.0 | –.(10)588 9829 883 | .(12)238 8265 393 |
| 12.5 | –.(10)348 0938 563 | .(12)130 2690 214 |
| 13.0 | –.(10)210 1044 796 | .(13)727 9739 432 |
| 13.5 | –.(10)129 2993 719 | .(13)415 9851 104 |
| 14.0 | –.(11)810 1043 368 | (13)242 6579 811 |
| 14 5 | – (11)516 0618 886 | .(13)144 2831 239 |
| 15.0 | –.(11)333 8664 755 | .(14)873 2925 919 |
| 15 5 | –.(11)219 1292 642 | .(14)537 4108 258 |
| 16.0 | –.(11)145 7726 865 | .(14)335 8817 661 |
| 16 5 | –.(12)982 0445 539 | .(14)212 9981 931 |
| 17.0 | –.(12)669 4697 055 | .(14)136 9274 099 |
| 17.5 | –.(12)461 4992 604 | .(15)891 6203 434 |
| 18.0 | –.(12)321 4946 024 | .(15)587 6588 627 |
| 18.5 | –.(12)226 1959 235 | .(15)391 7725 751 |
| 19.0 | –.(12)160 6464 099 | .(15)264 0206 484 |
| 19.5 | –.(12)115 1112 743 | .(15)179 7587 394 |
| 20.0 | –.(13)831 8162 819 | .(15)123 5841 333 |
| 20.5 | –.(13)605 9221 411 | .(16)857 5225 577 |
| 21.0 | –.(13)444 7507 764 | .(16)600 2657 904 |
| 21.5 | –.(13)328 8288 593 | .(16)423 7170 285 |
| 22.0 | –.(13)244 8106 616 | .(16)301 4909 626 |
| 22.5 | –.(13)183 4686 278 | .(16)216 1633 317 |
| 23.0 | –.(13)138 3685 504 | .(16)156 1179 618 |
| 23.5 | –.(13)104 9876 917 | .(16)113 5403 358 |
| 24.0 | –.(14)801 2235 814 | .(17)831 2774 587 |
| 24.5 | –.(14)614 8704 752 | .(17)612 5263 839 |
| 25.0 | –.(14)474 3701 124 | .(17)454 1098 277 |

# THE PRECISION OF THE WEIGHTED AVERAGE

*By*
H. MILICER GRUZEWSKA, PH. DR.,
*Warsaw, Poland.*

*Introduction.* We shall consider an infinite *universe* of el ments characterized by pairs of variable quantities $x_i$, $(i=1,2,3\cdots\cdots,\infty)$. Regarding the values of $y_i$ as the weigh to be assigned to the variates $x_i$ the weighted average of $x_i$ m be denoted by $x_y$, i.e.

$$x_y = \frac{x_1 y_1 + x_2 y_2 + x_3 y_3 \cdots \cdots}{y_1 + y_2 + y_3 \cdots \cdots \cdots} .$$

All possible samples, each of $N$ pairs of variates $x_i, y_i$ $(i=1,2,3\ldots,$ that can be selected from the universe constitute the sample popu lation.

Our problem is to obtain an expression for the probable pr cision of the weighted average $x_y$ according to certain hypothes concerning the selection of the pairs of variates in various sampl Professor Bowley discussed this problem in his paper on "Precisi of Measurement Attained in Sampling"[1] presented in Rome duri the Congress of Statistics 1925. In this paper Professor Bowl made no allowance for correlation between the variates $x_i$ and $y_i$ In the present paper I shall attempt to eliminate this restrictio

I am greatly indebted to Professor A L. Bowley for suggestio regarding the simplification of the proof of theorem II and for h general assistance in improving the form of this paper.

Let us suppose:

(a) the pairs of elements selected from the universe are indepe dent of each other,

---

[1]Cambridge 1925.

(b) the number of pairs in each sample is so large that $\frac{1}{N}$ may be neglected,

(c) the frequency surface $f(x,y)$ is normal, i.e. the probability $P_i$ that the particular pair $x_i, y_i$ will be selected is,

$$P_i = \frac{1}{\sigma_x \sigma_y 2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{x_i-x}{\sigma_x}\right)^2+\left(\frac{y_i-y}{\sigma_y}\right)^2-\frac{2r(x_i-x)(y_i-y)}{\sigma_x\sigma_y}\right]},$$

where $x, y, \sigma_x, \sigma_y$ and $r$ designate the parameters characterizing the surface,

(d) the à priori chance that the parameters of (c) are equal to given values may be defined by the function $F(x_y, y, \sigma_x, \sigma_y, r)$ where this function is integrable, can be expanded in Taylor's series and converges over the whole space.

Let the calculated characteristics of the sample be,

$X_y$ the weighted average of $x_i$ with $y_i$ as weights $(i=1,2,3\cdots,N)$
$Y$ the arithmetic average of the variates $y_i, (i=1,2,3,\cdots,N)$
$X$ the arithmetic average of the variates $x_i, ($ " " $)$
$S_x$ the standard deviation of the variates $x_i, ($ " " $)$
$S_y$ the standard deviation of the variates $y_i, ($ " " $)$
$R$ the coefficient of correlation between the variates $x_i$ and $y_i$ $(i=1,2,3\cdots,N)$ .

The expressions representing the most probable values of the weighted average and its standard deviation are independent whether the parameters of the universe are known or unknown. In Parts I, II, and III we shall consider the respective cases,

(a) when all parameters are unknown,
(b) all but $y$ are unknown,
(c) all but $y$ and $\sigma_y$ are unknown.

In Part IV we shall consider the generalized case of Part I when there are $k$ sets of elements, i.e. $x_i^t, y_i^t \begin{bmatrix} t=1,2\cdots\cdots k \\ i=1,2,3\cdots\cdots\infty \end{bmatrix}$

in the universe. In order to consider this case we shall, at the beginning of Part IV, slightly change the hypotheses and modify the above notation.

## PART I

### Case Where All Parameters Are Unknown

*Theorem (1.1).* If hypotheses (a) and (c) are satisfied and if $S_x S_y(1-R^2) \neq 0$ then, the most probable value of $x_y$ is $X_y$.

*Proof.* If $P_n$ denotes the probability of getting $N$ particular pairs of variates, then it follows from hypotheses (a) and (c) that,

(1)
$$P_n = \left(\frac{1}{\sigma_x \sigma_y 2\pi\sqrt{1-r^2}}\right)^N e^{-\frac{N}{2(1-r^2)}\left[\frac{S_x^2+(x-X)^2}{\sigma_x^2}+\frac{S_y^2+(y-Y)^2}{\sigma_y^2}-2r\,\frac{RS_xS_y+(x-X)(y-Y)}{\sigma_x\sigma_y}\right]}.$$

Taking the partial derivatives of $P_n$ with respect to $x, y, \sigma_x, \sigma_y$ and $r$, setting them equal to zero, and solving for $x, y, \sigma_x, \sigma_y$ and $r$, yields

(2)
$$\begin{cases} x = X, & \sigma_x = S_x, \\ y = Y, & \sigma_y = S_y, \end{cases} \quad r = R$$

hence $x = X$, $y = Y$, $\sigma_x = S_x$, $\sigma_y = S_y$ and $r = R$ will make $P_n$ a maximum, and the maximum value of $P_n$ is,

(3)
$$P_{max} = \left[\frac{1}{e S_x S_y 2\pi\sqrt{1-R^2}}\right]^N.$$

The weighted average $x_y$ and $X_y$ can be expressed in terms of $x, y, \sigma_x, \sigma_y, r$ and $X, Y, S_x, S_y, R$ respectively,

$$X_y = \sum_{i=1}^{N} (x_i \cdot y_i)/y_i \qquad \text{(by definition)}$$

$$= \frac{\frac{1}{N}\sum_{i=1}^{N} x_i\, y_i - XY + XY}{Y} \quad \text{(since } \frac{1}{N}\sum_{i=1}^{N} y_i = Y \text{ by definition)}$$

(4) hence,

$$X_y = \frac{R S_x S_y}{Y} + X \qquad \text{(since } \frac{1}{N}\sum_{i=1}^{N} x_i\, y_i - XY = R S_x S_y \text{)}$$

similarly.

$$x_y = \frac{r \sigma_x \sigma_y}{y} + x$$

*This proves theorem (1.1).*

*Theorem (1.2).* If all four hypotheses are satisfied and if $S_x S_y (1-R^2) \neq 0$ then the *à posteriori* probability $P$ that the sample came from the universe, the weighted average $x_y$ of which satisfies the inequality $\left| x_y - X_y \right| \leq \epsilon$ , can be expressed by,

(5)
$$P = \frac{1}{\sqrt{2\pi}\,\sigma} \int_0^{\epsilon} e^{\frac{t^2}{2\sigma^2}} dt \qquad \text{where}$$

$$\sigma = \frac{S_x}{\sqrt{N}} \sqrt{1 + \left(\frac{S_y}{Y}\right)^2 \left\{1 - R^2\left[1 - \left(\frac{S_y}{Y}\right)^2\right]\right\}}$$

*Proof.* It follows from (4) that,

$$x - X = x_y - X_y - \frac{r\sigma_x \sigma_y}{y} + \frac{RS_x S_y}{Y}$$

Substituting the above value of $(x-X)$ in (1) we shall have,

$$(6)\begin{cases} P_n = \left(\dfrac{1}{\sigma_x \sigma_y 2\pi\sqrt{1-r^2}}\right)^N e^{-\frac{N}{2(1-r^2)}W} & \text{where} \\[2ex] W = V^2 + \left(\dfrac{S_x}{\sigma_x}\right)^2 + \left(\dfrac{S_y}{\sigma_y}\right)^2 - 2rR\,\dfrac{S_x S_y}{\sigma_x \sigma_y} + (1-r^2)\left(\dfrac{y-Y}{\sigma_y}\right)^2 & \text{and} \\[2ex] V = -\dfrac{S_x}{\sigma_x}\left(\dfrac{x_y - X_y}{S_x} + \dfrac{RS_y}{Y}\right) + r\left(\dfrac{y-Y}{\sigma_y} + \dfrac{\sigma_y}{y}\right) \end{cases}$$

Let:
$$x_y - X_y = d'S_x, \quad \sigma_x - S_x = \lambda' S_x \qquad \rho = r - R$$
$$y - Y = d''S_y, \quad \sigma_y - S_y = \lambda'' S_y$$

then,

$$(7)\begin{cases} \dfrac{P_n}{P_{max}} = \left[\dfrac{e\sqrt{1-R^2}}{(1+\lambda')(1+\lambda'')\sqrt{1-(R+\rho)^2}}\right]^N e^{-\frac{N}{2[(1-(R+\rho)^2]}W_1} & \text{in which} \\[2ex] W_1 = V_1^2 + \dfrac{1}{(1+\lambda')^2} + \dfrac{1}{(1+\lambda'')^2} - \dfrac{2R(R+\rho)}{(1+\lambda')(1+\lambda'')} + \left[1-(R+\rho)^2\right]\left(\dfrac{d''}{1+\lambda''}\right)^2 & \text{and} \\[2ex] V_1 = -\dfrac{1}{(1+\lambda')}\left(d' + \dfrac{RS_y}{Y}\right) + (R+\rho)\left(\dfrac{d''}{1+\lambda''} + \dfrac{1+\lambda''}{\frac{Y}{S_y} + d''}\right). \end{cases}$$

Taking the logarithm of $\frac{P}{P_{max}}$ we shall have,

$$\frac{1}{N}\log\frac{P_{max}}{P_n} = A, \quad \text{where}$$

$$(8)\left\{\begin{aligned} &A = const. + \log(1+\lambda_1') \\ &+ \log(1+\lambda'') + \tfrac{1}{2}\log\left[1-(R+\rho)^2\right] + \frac{1}{2\left[1-(R+\rho)^2\right]}W \end{aligned}\right.$$

Expanding $A$ in term of the small quantities $\lambda', \lambda'', d', d'', \rho$ to second powers inclusive and letting $K = \frac{S_y}{Y}, \lambda = \lambda' + \lambda''$ we obtain

$$(8')\left\{\begin{aligned} &\frac{1}{N}\log\frac{P_{max}}{P_n} = A_1 + A_2 \qquad \text{where} \\ &A_1 = \frac{1}{2}\left(d'\frac{\partial A}{\partial d'} + \cdots \rho\frac{\partial A}{\partial \rho}\right)^{(2)} \qquad \text{or} \\ &2A_1 = const + \frac{4(\lambda' + \frac{\lambda}{2})}{(1-R^2)} \\ &+ \frac{1-R^2(1-K^2)}{(1-R^2)}\left[\lambda\, R\,\frac{Kd_1' - RK(1-K^2)d'' + (1-K^2)\rho}{(1-R^2)(1-K^2)}\right]^2 \\ &+ \frac{1-R^2+K^2(1-K^2)}{(1-R^2)(1-K^2)}\left[\frac{\rho}{1-R^2} - K\frac{d' - K(1-K^2)d''}{1-R^2+K^2(1+R^2)}\right]^2 \\ &+ \frac{1+K^2\left[1-R^2(1-K^2)\right]}{1-K^2+K^2(1+R^2)}\left[d'' - \frac{R(1-K^2)d_1'}{1+K^2\left[1-R^2(1-K^2)\right]}\right]^2 + \frac{d_1^2}{1+K^2\left[1-R^2(1-K^2)\right]} \\ &= A_1' + \frac{d_1^2}{1+K^2\left[1-R^2(1-K^2)\right]} \end{aligned}\right.$$

(We shall make use of the above substitution in the next paragraph.)

Therefore the probability of getting a particular set of $N$ pairs of variates can be expressed approximately by,

$$(9) \qquad P_n = a \text{ const times } e^{-NA_1}$$

Then it follows from hypothesis (d) and (8′) that the *à posteriori* probability $P$ that the sample came from the universe—the weighted average $x_y$ of which satisfies the inequality $[x_y - X_y] \leq \epsilon$, whatever the parameters $x, y, \sigma_x, \sigma_y$ and $r$ may be—is expressed by,

$$(10) \qquad P = \frac{\int_{X_y-\epsilon}^{X_y+\epsilon} \int_{-\infty}^{\infty} \cdots \int_{-1}^{1} F(x_y, \cdots r) e^{-N(A_1+A_2)} d_{x_y} \cdots dr}{\int_{-\infty}^{\infty} \cdots \int_{-1}^{1} F(x_y, \cdots r) e^{-N(A_1+A_2)} d_{x_y} \cdots dr}.$$

We may write,

$$(11) \qquad \begin{cases} F(x_y, y, \sigma_x, \sigma_y, r) e^{-N(A_1+A_2)} \\ \quad = F(x_y, y, \sigma_x, \sigma_y, r) e^{-N(A_1'/2+A_2) - \frac{\left(\frac{Nd_1^2}{1+K^2[1-R^2(1-K^2)]}\right)}{2}} \\ \quad = F_1(x_y, y, \sigma_x, \sigma_y, r) e^{-\frac{N(x_y-X_y)^2}{\sigma_0^2}} \end{cases}$$

where

$$\sigma_0 = \frac{S_x}{\sqrt{2}} \sqrt{1+K^2\left[1-R^2(1-K^2)\right]} \qquad \text{since } \left(\frac{(x_y - X_y)}{S_x}\right) = d_1$$

and

$$F_1(x_y, y, \sigma_x, \sigma_y, r) = F(x_y, \sigma_x, \sigma_y, r) e^{-N(A_1'/2+A_2)}$$

$$
(12)\begin{cases}
\text{Let,} \quad E = \sqrt{N}(xy - X_y) \\
\text{and} \\
\varphi\left(\frac{E}{\sqrt{N}}\right) = \int_{\infty}^{\infty}\int_{-1}^{1} F_1\left(\frac{E}{\sqrt{N}} + X_y, y \cdot r\right) dy \quad dr \\
\text{then,} \quad P = \int_{-\epsilon\sqrt{N}}^{\epsilon\sqrt{N}} \varphi\left(\frac{E}{\sqrt{N}}\right) e^{-\left(\frac{E}{\sigma_0}\right)^2} dE \Big/ \int_{\infty}^{\infty} \varphi\left(\frac{E}{\sqrt{N}}\right) e^{-\left(\frac{E}{\sigma_0}\right)} dE.
\end{cases}
$$

It follows from (8′), (11), and (12) and hypothesis (d) that $\varphi\left(\frac{E}{\sqrt{N}}\right)$ can be developed in Taylor's series for all values of $N$, hence,

$$
(13)\begin{cases}
\varphi\left(\frac{E}{\sqrt{N}}\right) = \varphi(0) + \frac{E}{\sqrt{N}}\varphi'(0) + \frac{1}{2!}\left(\frac{E}{\sqrt{N}}\right)^2 \varphi''(0) + \frac{1}{3!}\left(\frac{E}{\sqrt{N}}\right)^3 \varphi'''(0) + \\
\quad = \varphi(0) + \left(\frac{E}{\sqrt{N}}\right)\varphi'(0) + \frac{1}{N}\left[\frac{1}{2!} E^2 \varphi''(0) + \frac{E^3}{3!\sqrt{N}}\varphi'''(0)\right. \\
\quad \left. + \frac{E^4}{4!(\sqrt{N})^2}\varphi^{IV}(0) + \cdot \cdot\right] = \varphi(0) + \left(\frac{E}{\sqrt{N}}\right)\varphi'(0) + O\left(\frac{1}{N}\right).
\end{cases}
$$

Neglecting terms of order of $\left(\frac{1}{N}\right)$ we shall have,

$$
(14) \quad P = \int_{-\epsilon\sqrt{N}}^{\epsilon\sqrt{N}} \left[\varphi(0) + \frac{E}{\sqrt{N}}\varphi'(0)\right] e^{-\left(\frac{E}{\sigma_0}\right)^2} dE \Big/ \int_{-\infty}^{\infty} \left[\varphi(0) + \left(\frac{E}{\sqrt{N}}\right)\varphi'(0)\right] e^{-\left(\frac{E}{\sigma_0}\right)^2} dE
$$

but, $\displaystyle\int_{\epsilon\sqrt{N}}^{\epsilon\sqrt{N}} E e^{-\left(\frac{E}{\sigma_0}\right)^2} dE = \int_{-\infty}^{\infty} E e^{-\left(\frac{E}{\sigma_0}\right)^2} dE = 0$ (odd function)

and $\displaystyle\int_{-\infty}^{\infty} e^{-\left(\frac{E}{\sigma_0}\right)^2} dE = \sqrt{\pi}\,\sigma_0$

hence, $\displaystyle P = \frac{2}{\sqrt{\pi}\,\sigma_0}\int_0^{\epsilon\sqrt{N}} e^{-\left(\frac{E}{\sigma_0}\right)^2} dE$

*Let* $\sigma = \dfrac{\sigma_0}{\sqrt{2N}}$ and $E = \sqrt{N}t$

then $\displaystyle P = \frac{1}{\sqrt{2\pi\sigma}}\int_0^{\epsilon} e^{2\frac{t^2}{\sigma^2}} dt$ *This proves theorem (1.2)*

## PART II

### CASE WHERE $y$ IS CONSTANT[1]

*Theorem (2.1).* If hypotheses (a) and (c) are satisfied and if $S_x, S_y(1-R)^2 \neq 0$ then,

$$
(1) \begin{cases}
(1\text{a}) & \sigma_1 = S_x\sqrt{1+\overline{RK_1}^{\,2}} \\
(1.\text{b}) & \sigma_2 = S_y\sqrt{1+K_1^2} \\
(1\text{c}) & r_0 = R\sqrt{(1+K_1^2)/(1+\overline{RK_1}^{\,2})} \\
(1.\text{d}) & x_0 = X_y + \sigma_1 r_0\left[k + \frac{y}{Y}(k_1 - k)\right]
\end{cases}
$$

where $\sigma_1, \sigma_2, r_0$ and $x_0$ are the most probable values of $\sigma_x, \sigma_y, r$ and $x_y$ respectively, and,

$$
(2) \qquad K = \frac{S_y}{Y}, \quad K_1 = \frac{y-Y}{S_y}, \qquad k = \frac{\sigma_2}{Y}, \quad k_1 = \frac{y-Y}{\sigma_2}
$$

*Proof.* The probability of getting $N$ particular pairs of variates is given by (6) of Part I. Taking the partial derivatives of $P_n$ with respect to $\sigma_x$, $\sigma_y$, $r$ and $x_y$, and setting them equal to zero, we obtain,

$$
(3) \begin{cases}
(3.\text{a}) & \frac{\partial P_n}{\partial \sigma_x} = 2(1-r^2) + \sigma_x W'_{\sigma_x} = 0 \\
(3.\text{b}) & \frac{\partial P_n}{\partial \sigma_y} = 2(1-r^2) + \sigma_y W'_{\sigma_y} = 0 \\
(3.\text{c}) & \frac{\partial P_n}{\partial r} = 2(1-r^2) - 2rW - (1-r^2)W'_r = 0 \\
(3.\text{d}) & \frac{\partial P_n}{\partial x_y} = V = 0
\end{cases}
$$

[1]Case where all the parameters but $y$ are unknown.

where $W'_{\sigma_x}$, $W'_{\sigma_y}$ and $W'_r$ mean the partial derivatives of $W$

with respect to $\sigma_x$, $\sigma_y$ and $r$ respectively,

But,

$$\sigma_x W'_{\sigma_x} = 2V\sigma_x V'_{\sigma_x} - 2\left(\frac{S_x}{\sigma_x}\right)^2 + 2rR\,\frac{S_x S_y}{\sigma_x \sigma_y}$$

$$\sigma_y W'_{\sigma_y} = 2V\sigma_y V'_{\sigma_y} - 2\left(\frac{S_x}{\sigma_x}\right)^2 + 2rR\,\frac{S_x S_y}{\sigma_x \sigma_y} - 2(1-r^2)\left(\frac{y-Y}{\sigma_y}\right)^2$$

$$W'_r = 2VV'_r - 2R\,\frac{S_x S_y}{\sigma_x \sigma_y} - 2r\left(\frac{y-Y}{\sigma_y}\right)^2$$

since $V=0$, we obtain,

$$(3')\begin{cases} (3'.a) & (1-r^2) - \left(\dfrac{S_x}{\sigma_x}\right)^2 + rR\,\dfrac{S_x S_y}{\sigma_x \sigma_y} = 0 \\ (3'.b) & (1-r^2) - \left(\dfrac{S_y}{\sigma_y}\right)^2 + rR\,\dfrac{S_x S_y}{\sigma_x \sigma_y} - (1-r^2)\left(\dfrac{y-Y}{\sigma_y}\right)^2 = 0 \\ (3'.c) & r(1-r^2) - r\left[\left(\dfrac{S_x}{\sigma_x}\right)^2 + \left(\dfrac{S_y}{\sigma_y}\right)^2\right] + (1+r^2)R\left(\dfrac{S_x S_y}{\sigma_x \sigma_y}\right)^2 = 0 \end{cases}$$

Solving for $\sigma_x$, $\sigma_y$ and $r$ from (3') and making use of the substitutions from (2) we get the most probable value of $\sigma_x$, $\sigma_y$ and $r$,

$$(4)\begin{cases} (4.a) & \sigma_x = \sigma_1 = S_x\sqrt{1+R_1K_1^2} \\ (4.b) & \sigma_y = \sigma_2 = S_y\sqrt{1+K_1^2} \\ (4.c) & r = r_0 = R\sqrt{(1+K_1^2)/(1+RK_1^2)} \\ & \text{and from (3.d) we obtain,} \\ (4.d) & x_y = x_0 = X_y - \dfrac{RS_x S_y}{Y} + r_0\sigma_1(k+k_1) \end{cases}$$

Since $k_1 = \dfrac{k_1\sigma_2}{S_y}$ we get from (4.a), (4.b) and (4.c),

(4'.a) $\qquad S_x = \sigma_1\sqrt{1-r_0K_1^2}$

(4′.b) $$S_y = \sigma_2 \sqrt{1-k_1^2}$$

(4′.c) $$R = r_0 \sqrt{(1-k_1^2)/(1-\overline{r_0 k}^2)}$$

hence,

$$R S_x S_y = \sigma_1 \sigma_2 r_0 (1-k_1^2)$$

Substituting the above value of $R S_x S_y$ in (4.d) we obtain,

(4′.d) $$x_y = x_0 = X_y + \sigma_1 r_0 \left[ k + \frac{y}{Y}(k_1 - k) \right]$$

*This proves theorem (2.1)*

If we denote the maximum probability by $P_{max}$ then,

(5) $$P_{max} = \left[ \frac{1}{2\pi e \sigma_1 \sigma_2 \sqrt{1-r_0^2}} \right]^N$$

*Theorem (2 2).* If all four hypotheses are satisfied and if $S_x S_y (1-R^2) \neq 0$ then the *à posteriori* probability, $P$ that the sample came from the universe, the weighted average $x_y$ of which satisfies the inequality $|x_y - X_y| \leq \epsilon$, can be expressed by,

(6) $$\begin{cases} P = \dfrac{2}{\sqrt{2\pi}\,\sigma_Y} \displaystyle\int_0^{\epsilon} e^{-\frac{t^2}{2\sigma_Y^2}}\, dt \qquad \text{where} \\ \\ \sigma_Y = \dfrac{\sigma_1}{\sqrt{N}} \sqrt{k^2(1+r_0^2) + (1-r_0^2)\dfrac{(1+kk_1)^2}{(1-k_1)^2}} \end{cases}$$

*Proof.* Let $n_1 = 1-\overline{r_0 k_1}^2$, $n_2 = 1-k_1^2 > 0$ then by substituting the values of $S_x$, $S_y$ and $R$ from (4′.a), (4′.b) and (4′.c) we get,

[2] In this case the function $F(x_y, y, \sigma_x, \sigma_y, r)$ in (d) is $F(x_y, \sigma_x, \sigma_y, r)$

$$(7)\quad \begin{cases} P_n = \left(\dfrac{1}{\sigma_x \sigma_y 2\pi\sqrt{(1-r^2)}}\right)^{N} e^{-\frac{N}{2(1-r^2)}W} & \text{where} \\ W = V^2 + n_1\left(\dfrac{\sigma_1}{\sigma_x}\right)^2 + n_2\left(\dfrac{\sigma_2}{\sigma_y}\right)^2 - 2rr_o n_2 \dfrac{\sigma_1\sigma_2}{\sigma_x\sigma_y} + (1-r^2)\left(\dfrac{k_1\sigma_2}{\sigma_y}\right)^2 & \\ V = -\dfrac{x_y - x_o}{\sigma_x} - \dfrac{r_o\sigma_1}{\sigma_x}(k+k_1) + r\left(\dfrac{y-Y}{\sigma_y} + \dfrac{\sigma_y}{y}\right) & \text{and} \end{cases}$$

hence,

$$(8)\quad \frac{P_n}{P_{max}} = \left(e\,\frac{\sigma_1}{\sigma_x}\frac{\sigma_2}{\sigma_y}\sqrt{\frac{1-r_o^2}{1-r^2}}\right)^{N} e^{-\frac{N}{2(1-r^2)}W}$$

Taking the logarithm of $\frac{P}{P_{max}}$ and letting

$$\sigma_x = \sigma_1(1+\lambda'), \qquad r = (r_o+\rho)$$
$$\sigma_y = \sigma_2(1+\lambda''), \qquad x_y = x_o + \sigma_1 d'$$

we shall have,

$$(9)\quad \begin{aligned} &\frac{1}{N}\log\frac{P_{max}}{P_n} = A \qquad \text{where} \\ &A = Const + \log(1+\lambda') + \log(1+\lambda'') + \tfrac{1}{2}\log\left[1-(r_o+\rho)^2\right] + \frac{1}{2\left[1-(r_o+\rho)^2\right]}W. \end{aligned}$$

Expanding $A$ in terms of the small quantities $\lambda'$, $\lambda''$, $d'$ and $\rho$ we obtain,

$$(10)\quad \begin{cases} \dfrac{1}{N}\log\dfrac{P_{max}}{P_n} = const + A_1 + A_2 \quad \text{where} \\ A_1 = \dfrac{1}{2(1-r_o^2)}\Big\{d'^2 + \left[2-r_o^2+r_o^2k(k+2k_1)\right]\lambda'^2 + \left[2-r_o^2+r_o^2k(k-2k_1)\right]\lambda''^2 \\ \quad + \left[\dfrac{1+r_o^2}{1-r_o^2} + k(k+2k_1)\right]\rho^2 - 2r_o(k+k_1)d'd_r' - 2r_o(k+k_1)d'\lambda'' \\ \quad - 2(k+k_1)d'\rho - 2r_o^2(1-k_1^2)\lambda'\lambda'' \\ \quad - 2r_o\left[1-k(k+2k_1)\right]\lambda'\rho - 2r_o(1-k_1^2)\lambda''\rho\Big\} \quad \text{and} \\ A_2 = \displaystyle\sum_{n=3}^{\infty}\frac{1}{n!}\left(\lambda'\frac{\partial A}{\partial\lambda'} + \lambda''\frac{\partial A}{\partial\lambda''} + \frac{d'\partial A}{\partial d'} + \frac{\rho\,\partial A}{\partial\rho}\right)^{(n)} \end{cases}$$

The expression representing the value of $A_1$ is quadratic in form in terms of the variables $\lambda'$, $\lambda''$, $d'$ and $\rho$ where all the coefficients are positive,

$$
\begin{aligned}
A_1 = {} & \frac{2-r_0^2+r_0^2k(k+2k_1)}{2(1-r_0^2)}\left\{\lambda'-r_0\frac{(k_1+k)d'+r_0(1-k^2)d''+[1-k(k+2k_1)]\rho}{2-r_0^2+r_0^2k(k+2k_1)}\right\}^2 \\
(11) \quad & + \frac{4[1-r_0^2+r_0^2k(k+2k_1)]}{2(1-r_0^2)[2-r_0^2+r_0^2k(k+2k_1)]}\left\{\lambda''-r_0\frac{[k(1-r_0^2k_1^2)-k_1(1-r_0^2)]d'+(1-k^2)\rho}{2[1-r_0^2+r_0^2k^2(1-r_0^2k_1^2)]}\right\}^2 \\
& + \frac{(1-r_0^2)(1+kk_1)^2+(1+r_0^2)k^2(1-k_1^2)}{2[1-r_0^2+r_0^2k(k+2k_1)]}\left\{\frac{\rho}{1-r_0^2}-\frac{[k+k_1-r_0^2k_1(1+kk_1)]d'}{(1-r_0^2)(1+kk_1)^2+(1+r_0^2)k^2(1-k_1^2)}\right\}^2 \\
& + \frac{d'^2(1-k_1^2)}{2[(1-r_0^2)(1+kk_1)^2+(1+r_0^2)k^2(1-k_1^2)]} \; .
\end{aligned}
$$

For the rest of the proof of this theorem we proceed as in Part I and can obtain,

$$
(12) \quad \sigma_Y = \frac{\sigma_1}{\sqrt{N}}\sqrt{k(1+r_0^2)+(1-r_0^2)\frac{(1+kk_1)^2}{(1-k_1)^2}} \; .
$$

*Notice that if* $y = Y$ *then,*

$$K_1 = k_1 = 0\,, \quad \sigma_1 = S_x\,, \quad \sigma_2 = S_x\,, \quad r_0 = R \quad \text{and} \quad x_0 = X_y$$

$$
(13) \qquad \sigma_Y = \sigma_{Y'} = \frac{S_x}{\sqrt{N}}\sqrt{1-R^2+k^2(1+R^2)}
$$

*hence* $\sigma_{Y'} < \sigma$ *if* $R \neq 0$ *where* $\sigma$ *is given by (5) Part I.*

## PART III

### Case Where $y$ and $\sigma_y$ are Constants[8]

*Theorem (3.1).* If hypotheses (a) and (c) are satisfied and if $S_x S_y (1-R)^2 \neq 0$ then,

[8]Case where all the parameters but $y$, $\sigma_y$ are unknown.

$$(1)\begin{cases} (1.a) & \sigma_1 = \dfrac{S_x}{S_y}\sqrt{R^2\sigma_y^2+(1-R^2)S_y^{\,2}} \\ (1.b) & r_0 = R\sigma_y \Big/ \sqrt{R^2\sigma_y^{\,2}+(1-R^2)S_y^{\,2}} \\ (1.c) & x_0 = X_Y - \dfrac{RS_x S_y}{Y} + \dfrac{RS_x S_y}{S_y}\left(\dfrac{y-Y}{\sigma_y}+\dfrac{\sigma_y}{y}\right) = X + \dfrac{RS_x C}{a} \end{cases}$$

where $\sigma_1$, $r_0$ and $x_0$ are the most probable values of $\sigma_x$, $r$ and $x_y$ respectively and

$$\frac{S_y}{Y} = K; \quad \frac{S_y}{\sigma_y} = a; \quad \frac{y-Y}{\sigma_y} + \frac{\sigma_y}{y} = c$$

*Theorem (3.2).* If all four hypotheses[4] are satisfied and if $S_x S_y(1-R^2) \neq 0$ then the *à posteriori* probability $P$ that the sample came from the universe, the weighted average $x_y$ of which satisfies the inequality $|x_y - X_y| \leq \epsilon$, can be expressed by.

$$(2) \qquad P = \frac{2}{2\pi\sigma_{Y_1}}\int_0^{\epsilon} e^{-\frac{t^2}{2\sigma_{Y_1}^2}}\,dt \quad \text{where}$$

$$\sigma_{Y_1} = \frac{\sigma_1}{\sqrt{N}}\sqrt{(1-r_0^{\,2})\left[1+\left(\frac{c}{a}\right)\right]}$$

*Notice that if* $y - Y$ *and* $\sigma_y = S_x$ *then,*

$$(3)\begin{cases} \sigma_1 = S_x, \quad r_0 = R; \quad x_0 = X_y & \textit{and} \\ \sigma_{Y_1} = \sigma_{Y_1'} = \dfrac{S_x}{\sqrt{N}}\sqrt{(1-R^2)(1+k^2)} \end{cases}$$

*hence* $\sigma_{Y_1} < \sigma_{Y'}$ *if* $R \neq 0$
*where* $\sigma_{Y'}$ *is given by (12) Part II.*

As the proofs of theorems (3.1) and (3.2) do not differ from the proofs of theorems (2.1) and (2.2.), we shall omit them *

[4]In this case the function $F(x_y, y, \sigma_x, \sigma_y, r)$ in (d) is $F(x_y, \sigma_x, r)$

*Part I and II were presented in Wilno during the II Assembly of Polish Mathematicians.

## PART IV

In this Part we shall consider the generalized case of Part I where there are $k$ sets of elements characterized by pairs of variable quantities, $x_i^{\ell}, y_i^{\ell} \begin{Bmatrix} \ell = 1, 2, 3, \cdots\cdots, k \\ i = 1, 2, 3, \cdots\cdots, \infty \end{Bmatrix}$

Let, $$x = \frac{\sum_1^k x_y^{\ell} A_{\ell}}{\sum_1^k A_{\ell}} = \sum_1^k x_y^{\ell} \frac{A_{\ell}}{A}$$

where $x_y^{\ell}$ is the weighted average of the variates $x_i^{\ell}$, with $y_i^{\ell}$ as weights, and $A_{\ell}$ the sum of these weights. Our problem is to obtain an expression for the probable precision of the quantity $x$ according to certain hypotheses.

We shall replace hypothesis (b) of the introduction by hypotheses (b′) and (b″) where,

(b′) the number ($N = N_1 + N_2 + N_3 + \cdots\cdots + N_k$) of pairs in each sample is so large that $\frac{1}{N}$ may be neglected,

(b″) each of the numbers $N_{\ell}$ ($\ell = 1, 2, 3, \cdots, k$) of pairs from separate sets is so large that $\frac{N_{\ell}}{N}$ has a significant value, i.e.,

$$\frac{N_{\ell}}{N} \geq \omega_0 > 0$$

Let us replace in hypothesis (c) $P_i$ by $P_i^{\ell}$ ($\ell = 1, 2, 3, \cdots, k$) and $x, y, \sigma_x, \sigma_y, r$ by $x^{\ell}, y^{\ell}, \sigma_x^{\ell}, \sigma_y^{\ell}, r_{\ell}$ and refer to the corresponding general hypothesis by (c′). Likewise if in hypothesis (d) we replace $F(x_y, y, \sigma_x, \sigma_y, r$ by $F(x_y^{\ell}, y^{\ell}, \sigma_x^{\ell}, \sigma_y^{\ell}, r_{\ell})$ we obtain the generalized hypothesis (d′).

We shall denote the calculated characteristics of the sample by

$$X_y^\ell, X^\ell, Y^\ell, S_x^\ell, S_y^\ell, R_\ell \; (i=1,2,3,\cdots; N_\ell)$$

corresponding to the values $X_y, X, Y, S_x, S_y, R$ as defined in the introduction page 197.

*Theorem* (4.1). If hypotheses (a) and (c′) are satisfied and if $(1-R)S_x^\ell S_y^\ell \neq 0$ then the most probable value of $x$ is $X$ where,

$$X = \sum_1^k X_y^\ell \frac{A_\ell}{A}$$

*Proof**. Let $P_n$ be the probability of getting a given set of $N$ pairs of variates $x_i^\ell, y_i^\ell$, then it follows from hypotheses (a) and (c′) that,

$$P_n = \prod_1^k \frac{e^{-\frac{N_\ell}{2(1-r_\ell^2)} W_\ell}}{\left(2\pi\sigma_x^2\sigma_y^2\sqrt{1-r_\ell^2}\right)^{N_\ell}}$$ , where

(1)

$$W_\ell = V_\ell^2 + \left(\frac{S_x^\ell}{\sigma_x^\ell}\right)^2 + \left(\frac{S_y^\ell}{\sigma_y^\ell}\right)^2 - 2r_\ell R_\ell \frac{S_x^\ell S_y^\ell}{\sigma_x^\ell \sigma_y^\ell} + (1-r_\ell^2)\left(\frac{y^\ell - Y^\ell}{\sigma_y^2}\right)^2$$ , and

$$V_\ell = -\frac{S_x^\ell}{\sigma_x^\ell}\left(\frac{x_y^\ell - x_y^\ell}{S_x^\ell} + \frac{R_\ell S_y^\ell}{Y^\ell}\right) + r_\ell\left(\frac{y^\ell - Y^\ell}{\sigma_y^\ell} + \frac{\sigma_y^\ell}{y^\ell}\right)$$

*The proofs of the theorems (4.1) and (4.2) shall be given in very abbreviated form as the method of proofs of these theorems does not differ from the proofs of theorem (2.1) and (2.2) of Part I.

Let,

$$
\text{(2)}\quad \left.\begin{array}{l}
x - X = D, \quad x_y^{\ell} - X_y^{\ell} = S_x^{\ell} d_{\ell}; \\[2ex]
y^{\ell} - Y^{\ell} = S_y^{\ell} \delta_{\ell}; \quad \sigma_x^{\ell} = S_x^{\ell}(1+\lambda_{\ell}^{1}); \\[2ex]
\sigma_y^{\ell} = S_y^{\ell}(1+\lambda_{\ell}^{\prime\prime}); \quad r_{\ell} = R_{\ell}(1+\rho_{\ell}) \\[2ex]
S_x^{\ell} \dfrac{A_{\ell}}{A} = \alpha_{\ell}; \quad \dfrac{S_y}{Y_{\ell}} = K_{\ell}
\end{array}\right\} \quad \ell = 1, 2, \quad k
$$

then,

$$
\text{(3)}\qquad D = \sum_{1}^{k} \alpha_{\ell}\, d_{\ell}
$$

and we can also express the unknown quantity $d_{\ell} = (\ell = 1, 2, \cdot\ , k)$

in terms of $D$ and the independent variable $\eta_{\ell}\ (\ell = 1, 2, 3\ \ \cdot\cdot, k-1)$
as follows,

$$
\text{(4)}\quad \begin{array}{l}
d_{\ell} = \dfrac{1}{\alpha_{\ell}}\left(\dfrac{D}{k} - \eta_{\ell}\right); \quad \ell = 1, 2, \ldots, k-1; \\[2ex]
d_{K} = \dfrac{1}{\alpha_{K}}\left(\dfrac{D}{K} + \sum_{1}^{k-1} \eta_{\ell}\right)
\end{array}
$$

Hence it follows from (I) that,

$$P_n = \prod_{\ell=1}^{k} \frac{e^{-\frac{N_\ell W_\ell}{2[1-R_1^2(1+\rho_\ell)^2]}}}{\left[(2\pi S_x^\ell S_y^\ell (1+\lambda'_\ell)(1+\lambda''_\ell)\sqrt{1-R_\ell^2(1+\rho_\ell)^2}\right]^{N_\ell}}$$ where

(5) $$W_\ell = V_\ell^2 + \frac{1}{(1+\lambda'_\ell)^2} + \frac{1}{(1+\lambda''_\ell)^2} - 2R_\ell^2 \frac{1+\rho_\ell}{(1+\lambda'_\ell)(1+\lambda''_\ell)} + \left[1-R_1^2(1+\rho_\ell)^2\right]\left(\frac{\delta_\ell}{1+\lambda''_\ell}\right)^2$$

and

$$V_\ell = -\frac{1}{1+\lambda'_\ell}(d_\ell + R_\ell K_\ell) + R_\ell(1+\rho_2)\left(\frac{\delta_\ell}{1+\lambda''_\ell} + \frac{1+\lambda''_\ell}{\delta_\ell + K_\ell}\right)$$

where $d_\ell$ are to be found from the equations (3) and $(\ell=1,2\cdots k)$

Taking the partial derivitives of $P_n$ with respect to $D$ and $\eta_\ell$ we obtain,

(6)
$$\frac{\partial P_n}{\partial D} = \frac{1}{k}\sum_1^k \frac{1}{\alpha_k}\cdot\frac{\partial P_n}{\partial d_\ell}$$

$$\frac{\partial P_n}{\partial \eta_\ell} = \frac{1}{\alpha_k}\frac{\partial P_n}{\partial d_k} - \frac{1}{\alpha_\ell}\cdot\frac{\partial P_n}{\partial d_\ell}\cdot \quad \ell = 1,2, \quad \cdot\cdot k-1$$

It can be easily vertified that $\frac{\partial P_n}{\partial D} = \frac{\partial P_n}{\partial \eta_\ell} = 0$ if and only if

$$\frac{\partial P_n}{\partial d_\ell} = 0$$

The probability $P_n$ treated as the function of variables

$D, \eta_1, \cdots \eta_{k-1}, \delta_1, \cdots \delta_k, \lambda'_1, \cdots \lambda'_k, \lambda''_2 \cdots \lambda''_k, \rho_1 \cdots \rho_k$, is a maximum when,

$$D = \eta_1 = \cdot\ \eta_{k-1} = \delta_\ell = \lambda'_\ell = \lambda^2_\ell = \rho_\ell = 0, (\ell = 1,2, \ \cdot\cdot k)$$

*This proves theorem (4.1).*

*Theorem* (4.2). If all hypotheses are satisfied and if then the *a' posteriori* probability that the sample came from the universe, the quantity $x$ of which satisfies the inequality $|x-X| \leq \epsilon$ may be expressed by,

$$P = \frac{2}{2\pi\sigma}\int_0^{\epsilon} e^{-\frac{t^2}{2\sigma^2}}\,dt,$$ where

(7) $$\sigma = \sqrt{\sum_1^k \left(s_x^\ell \frac{A_\ell}{A}\right)^2 \frac{\Phi_\ell}{N_\ell}},$$ and

$$\Phi_\ell = 1 + \left(\frac{s_y^\ell}{Y^\ell}\right)^2 \left\{1 - R_\ell^2\left[1 - \left(\frac{s_y^\ell}{Y^\ell}\right)^2\right]\right\}, (\ell = 1, 2, \cdots k)$$

*Proof.* Let $P_{max}$ denote the maximum probability, then it follows from (6) that,

(8) $$P_{max} = e^{-N} \prod_1^k \frac{1}{\left(2\pi s_x^\ell s_y^\ell \sqrt{1-R_\ell^2}\right)^{N_\ell}}$$ and

$$\frac{P_n}{P_{max}} = e^N \prod_1^k \left[\frac{\sqrt{1-R_\ell^2}}{(1+\lambda_\ell')(1+\lambda_\ell'')\sqrt{1-R_\ell^2(1+\rho_\ell)^2}}\right]^{N_\ell} e^{-\frac{n_\ell W_\ell}{2[1-R_\ell^2(1+\rho_\ell)^2]}}$$

$$= e^N \prod_1^k \left[\frac{\sqrt{1-R_\ell^2}}{(1+\lambda_\ell')(1+\lambda_\ell'')\sqrt{1-R_\ell^2(1+\rho_i)^2}}\right]^{Nw_\ell} e^{-\frac{Nw_\ell W_\ell}{2[1-R_\ell^2(1+\rho_\ell)^2]}}$$

where the value of $W_\ell$ given by (5) and $w_\ell = \frac{N_\ell}{n}$

As in Part I or Part II if we expand the $\log \frac{P_n}{P_{max}}$ in terms of $D, \eta_1, \cdots \eta_{k-1}, \delta_1 \cdots \delta_k, \lambda_1' \cdots \lambda_1^k, \lambda_2' \cdots \lambda_2^k, \rho_1 \cdots \rho_k$, the first term that does not vanish is quadratic in form in terms of the variables,

$$D, \eta_1, \cdots \eta_{k-1}; \delta_1, \cdots \delta_k; \lambda_1', \cdots \lambda_1^k; \lambda_2', \cdots \lambda_2^k; \rho_1, \cdots \rho_k;$$

and this in turn by linear transformation can be expressed as,

$$
(9)\begin{cases}
N(C_1 D^2 + C_2 \bar{\eta}_1^2 + \cdots + C_{5k} \bar{\rho}_k^2); & \rho_l > 0\ (l = 1,2,3,4,5k) \\[2ex]
C_1 = \dfrac{1}{\sum_1^k \dfrac{\alpha_l^2 \varphi_l}{\omega_l}} & \text{when} \\[2ex]
\varphi_l = 1 + K_l^2 \left[ 1 - R_l^2 (1 - k_l^2 \right] & \text{and}
\end{cases}
$$

To complete the proof we proceed as in Part I.

H. Milicer-Gruzewska

# ON CERTAIN RELATIONSHIPS BETWEEN $\beta_1$ AND $\beta_2$ FOR THE POINT BINOMIAL.*

*By* MARGARET MERRELL

## 1. *Introduction.*

The extensive literature on the point binomial covers studies on a variety of its properties, apropos of its use as a discrete probability function and of its approximate representation by certain continuous curves. Investigations on such properties as the sum of its terms within specified limits, the ratio of its ordinates, and the slope of chords connecting successive ordinates have thrown light on the characteristics of the point binomial and have suggested various continuous functions as substitutions for the binomial expansion.

Prominent among such studies have been those dealing with the moments of the binomial and it is with these properties that the present paper is concerned. The first four moments have been used by Pearson[1] as a means of fitting a point binomial to observed data and he has pointed out[2] that these moments expressed in terms of $\beta_1$ and $\beta_2$ approach the corresponding moments of the normal curve as $n$ becomes indefinitely large. Other papers that especially concern the following discussion are one by Student[3] in which he discussed the relationship between $\beta_1$ and $\beta_2$ for the point binomial and the Poisson exponential series, and one by Lucy Whitaker,[4] in which the range of $\beta_1$ and $\beta_2$ and certain relation-

---

*Paper No. 175 from the Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Md.

[1]Pearson, K. Skew variation in homogeneous material. Phil Trans. Vol. *186 A*, (1895), pp. 343-414.

[2]Pearson, K. On the curves which are most suitable for describing the frequency of random samples of a population. Biometrika, Vol. *5* (1906) pp. 172-175.

[3]Student. On the error of counting with a haemacytometer. Biometrika, Vol *5* (1906), pp. 351-360

[4]Whitaker, Lucy. On the Poisson law of small numbers. Biometrika Vol. *10* (1914), pp. 36-71

ships between the moments and the constants of the point binomial were discussed.

The present note will give some additional relationships between the third and fourth moments of the point binomial, in terms of $\beta_1$ and $\beta_2$, and will discuss these relationships in connection with their bearing on the association of the point binomial and the normal curve. The point binomial, $(p+q)^n$, has of course two constants $p$ and $n$ which completely determine its characteristics. Certain of these properties are closely connected with $\beta_1$ and $\beta_2$, and it is therefore of interest to see how $\beta_1$ and $\beta_2$ change as $p$ and $n$ take on different values In order to see the effect of varying each of the constants, the relationship between $\beta_1$ and $\beta_2$ will be determined for varying values of $n$ when $p$ is held constant, and for varying values of $p$ when $n$ is held constant. In addition to these relationships, it is possible to see how the $\beta$'s are related when both $p$ and $n$ are allowed to vary while certain functions of these parameters are held constant. In the following discussion the relationship between $\beta_1$ and $\beta_2$ will be considered for the cases where the mean, $np$, is held constant, and where the square of the standard deviation, $npq$, is held constant, $n$ and $p$ being variable.

The moments of the point binomial $(p+q)^n$ are:

$$mean = np$$

$$\mu_2 = npq$$

$$\mu_3 = npq(q-p)$$

$$\mu_4 = npq\left[1+3pq(n-2)\right]$$

These moments lead to the following values of the $\beta$'s:

(1) $$\beta_1 = \frac{(q-p)^2}{npq},$$

(2) $$\beta_2 = \frac{1+3pq(n-2)}{npq}.$$

Although the point binomial is ordinarily applied only to the case where $p$, $q$, and $n$ are positive, it should be noted that thes formulae for $\beta_1$ and $\beta_2$ are not limited to this case. The onl limitation on these constants is that $p + q = 1$

2. *The relationship between $\beta_1$ and $\beta_2$ for constant values of $p$*

If we eliminate $n$ between (1) and (2) we obtain an equatio relating $\beta_1$, $\beta_2$, and $p$, $n$ being unspecified. This equation is

$$\beta_2 = 3 + \frac{1-6pq}{(q-p)^2}\beta_1 \tag{3}$$

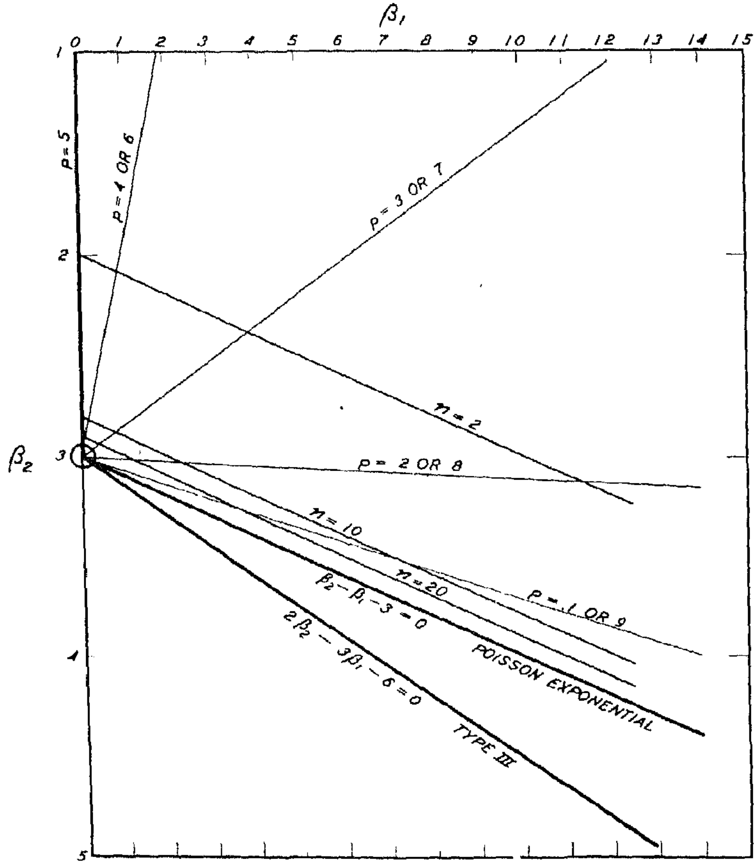



FIG 1 THE RELATIONSHIP BETWEEN $\beta_1$ AND $\beta_2$ FOR POINT BINOMIALS HA CONSTANT VALUES OF $p$ AND CONSTANT VALUES OF $n$

or fixed values of $p$, this represents a family of straight lines ith slope $\frac{1-6pq}{(q-p)^2}$, all passing through the position $\beta_1=0, \beta_2=3$. e position of the $\beta$'s for the normal curve. Figure 1 shows a 'oup of these lines on the $\beta_1 \beta_2$ plane, for the values of $p$ indited. The various positions on these lines represent $\beta_1$ and $\beta_2$ ›r point binomials having the specified $p$'s and varying $n$'s. Only ose values of $p$ which are between $0$ and $1$ are included in this agram, and these only for positive values of $n$, since these values ›ver all the ordinary probability problems. From (1) it can be en that for such values of $p$ and $n$, $\beta_1$ is positive The $\beta$'s for nomials having parameters outside these limits will be discussed a later paragraph.

The range through which these lines can swing can be deterined by substituting the limiting values of $p$ in equation (3). he values of $p$ to be used in determining this range are $0$ and $.5$, ›t $0$ and $1$. This follows from the fact that, since $p$ and $q$ are terchangeable in the point binomial, and the slope of the lines ven by (3) is symmetrical in $p$ and $q$, the lines obtained for $p$ :tween $.5$ and $1$ would be identical with the lines for the comement of $p$, $1-p$, between $0$ and $.5$. Thus any particular line presents point binomials for two complementary values of $p$.

If we substitute the value $.5$ for $p$ in (3), the equation :comes

$$\beta_1 = 0 \tag{4}$$

, we would expect, since point binomials having a $p$ or $.5$ are 'mmetrical. If $p=0$, equation (3) becomes

$$\beta_2 - \beta_1 - 3 = 0. \tag{5}$$

he latter line is identical with the line giving the relationship beveen $\beta_1$ and $\beta_2$ in the Poisson exponential series. This is in ırmony with the derivation of the Poisson exponential as the miting case of the point binomial as $p$ tends to $0$, and $n$ tends

to $\infty$, $np$ being finite. If we denote the mean of this series by $m$, the moments are[5]

$$\mu_2 = m$$

$$\mu_3 = m$$

$$\mu_4 = 3m^2 + m$$

and

$$\beta_1 = \frac{1}{m}$$

$$\beta_2 = 3 + \frac{1}{m} \, .$$

We have thus the equation relating $\beta_1$ and $\beta_2$ as given by (5).

The radiating lines giving the $\beta$'s for point binomials having values of $p$ between $0$ and $1$, will therefore lie in the range between the vertical, $\beta_1 = 0$, and the Poisson exponential line. This range, which is indicated in figure 1, was pointed out by Lucy Whitaker[6] in the paper mentioned above. The Type III line is included in this graph to indicate that portion of the $\beta_1 \beta_2$ plane covered by this family of lines. It is of interest to note that $\beta_1$ and $\beta_2$ for skew binomials do not approach the $\beta$'s of the Type III curve, although Pearson[7] has shown that in an important slope property the skew binomial polygon and the Type III curve follow the same law.

*3. The relationship between $\beta_1$ and $\beta_2$ for constant values of $n$.*

The $\beta_1 \beta_2$ equations for constant values of $p$ have been expressed as continuous straight lines, but only certain positions on these lines pertain to binomials having integral values of $n$. These points are determined by the intersection of these lines with the curve relating $\beta_1$, $\beta_2$, and $n$, when $n$ is held constant at integral

[5]Student. Loc cit. p. 353.

[6]Whitaker, Lucy. Loc. cit p. 37

[7]Pearson, K. Skew variation in homogeneous material. Loc. cit. p. 357.

values. The equation of this curve given by eliminating $p$ and $q$ between (1) and (2) is:

$$\beta_2 - \beta_1 - \frac{3n-2}{n} = 0. \tag{6}$$

This is a family of straight lines parallel to the Poisson exponential line, a specified value of $n$ determining a particular line. The intersection of any of these lines with any $p$ line determines the $\beta$'s for the point binomial of specified $p$ and $n$. Figure 1 gives the graph of three such lines for $n=2$, $n=10$, and $n=20$. From this graph it can be seen that even with an $n$ as small as $20$, $\beta_1$ and $\beta_2$ for the symmetrical and slightly skew binomials are not far from the position of the $\beta$'s of the normal curve, but for the highly skew binomials, they are quite far from this position. This is in agreement with the fact that the more skew the binomial, the larger the $n$ required to make the normal curve a good substitute for the binomial expansion.

It is evident from this graph that as $n$ is fixed at increasingly large values, $\beta_1$ and $\beta_2$ for the point binomials of different $p$'s converge quite rapidly toward the normal position. The limit of equation (6), as $n$ becomes indefinitely large, is equation (5), that is, the line giving $\beta_1$ and $\beta_2$ for the Poisson exponential. This is to be expected, considering the conditions under which the point binomial approaches the Poisson exponential. For this limiting case the $\beta_1, \beta_2$ line for constant $n$ crosses all the radiating $p$ lines at $0, 3$, except the line for $p=0$, with which it coincides throughout Thus for all values of $p$, except $0$, $\beta_1$ and $\beta_2$ for the point binomial agree with the corresponding moments of the normal curve, as $n$ becomes indefinitely great.

4. *The relationship between $\beta_1$ and $\beta_2$ for constant values of $np$.*

In judging how adequate the size of a particular sample is, for a specified value of $p$, we frequently make our estimate in terms, not of $n$, but of the mean value, $np$. This is with the thought that

we are in approximately as good a position with a $p$ of $.1$ and an $n$ of 50, for example, as with a $p$ of .01 and an $n$ of 500, since the expected number is the same in both cases. For instance, our knowledge of a penny from $10$ tosses is about as complete as that of a dice from $30$ tosses. To study this question from the moments of the binomial we can determine the curve relating the $\beta$'s for binomials of constant $np$, by replacing $np$ by $m$ in equations (1) and (2), and eliminating the remaining $p$'s and $q$'s between these two equations. This gives the equation

$$(7)\qquad \begin{aligned} &3m^2\beta_1^2 - 5m^2\beta_1\beta_2 + 2m^2\beta_2^2 + 3m(5m-2)\beta_1 \\ &\quad - 4m(3m-1)\beta_2 + 2(3m-1)^2 = 0 \end{aligned}$$

This is the equation of a hyperbola with asymptotes

$$(8)\qquad 2\beta_2 - 3\beta_1 - 6 = 0$$

$$(9)\qquad \beta_2 - \beta_1 - \frac{3m-2}{m} = 0.$$

The substitution of any particular value of $m$ in equation (7) will give the curve of $\beta$'s for all binomials having this specified mean value. Its intersection with the radiating lines of constant $p$ gives $\beta_1$ and $\beta_2$ for the point binomial having the specified $p$ and mean value. Figure 2 shows the two hyperbolas for which $m = 2$ and 10 respectively.

Turning to the asymptotes of the hyperbola, it will be seen that only one of them varies with $m$. Thus the various hyperbolas obtained by substituting different values of $m$ in (7) will all be asymptotic to the same line (8). The centers of all the hyperbolas will therefore lie on this line, which, it will be noted, is the Type III line. The other asymptote is parallel to the Poisson exponential line and a comparison of its equation, (9), with equation (6) shows that it represents the same relationship between $\beta_1$, $\beta_2$, and $m$, as that previously derived between $\beta_1$, $\beta_2$, and $n$. This asymptote is therefore the particular line in the family of $n$ lines for which $n = m$ or $np$. Since any point on the hyperbola is the

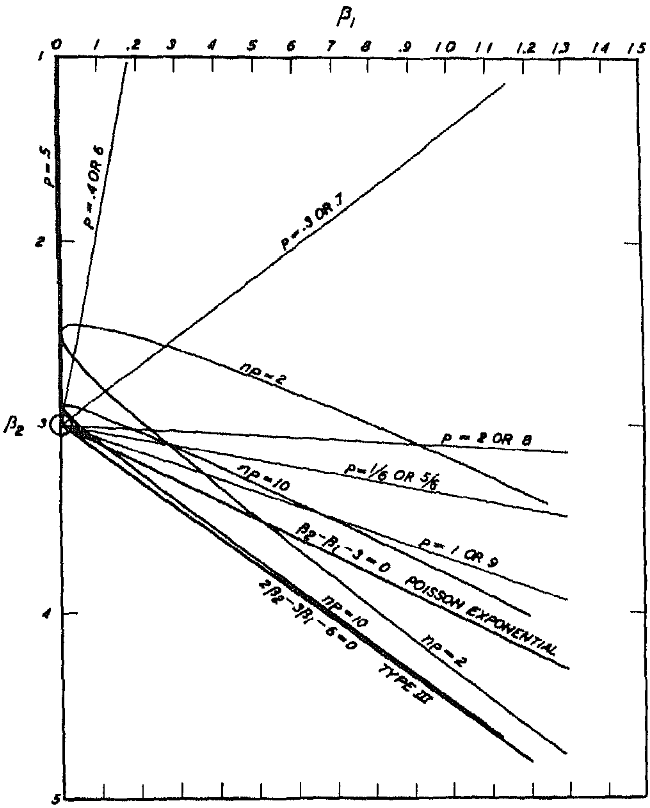



FIG 2 THE RELATIONSHIP BETWEEN $\beta_1$ AND $\beta_2$ FOR POINT BINOMIALS HAVING CONSTANT VALUES OF $np$

intersection of $n$ and $p$ lines such that $np$ has the constant value $m$, it follows that points farther and farther out on the hyperbola represent the intersection of lines having values of $n$ closer and closer to the mean, $np$, and values of $p$ approaching the value $1$. If we consider this asymptote for hyperbolas of increasing values of $m$ we see that it approaches the Poisson exponential line, and in the limiting case, as $m$ becomes indefinitely large, the hyperbola degenerates into the two lines which represent the $\beta$'s for the Poisson exponential series, and the Type III curve. These limits will be further discussed in a later paragraph.

All of the hyperbolas in the family represented by equation (7) are tangent to the line $\beta_1 = 0$, which is the value for binomials of $p = .5$. From figure 2 it can be seen that the lines giving the $\beta$'s for the other values of $p$ are crossed twice by each hyperbola. This is due to the fact that each line represents the $\beta$'s for two complementary values of $p$, and for such values of $p$ the same mean value would result from two different values of $n$, such that

$$m = np = n'(1-p)$$

For example, the binomials $(.2 + .8)^{20}$ and $(.8 + .2)^{5}$ have $p$ values that lie on the same straight line and mean values that lie on the same hyperbola It is thus obvious why the hyperbola must be tangent to the line for $p = .5$, since in this case the two complementary values of $p$ are equal and there can be only one value of $n$ which will produce a given $m$.

From the discussion of lines relating $\beta_1$ and $\beta_2$ for constant $n$ values, it is evident that of the two crossings of any $p$ line, the one nearer the Gaussian position is for the point binomial with the larger $n$ and therefore for the smaller of the two complementary $p$ values. Furthermore, through the point of intersection of two hyperbolas, only one $n$ line and one $p$ line will pass, and the $\beta$'s thus determined are for two binomials, one having the smaller mean and the smaller $p$, and the other the larger mean and $p$, both having the same $n$. For example, the two hyperbolas given in figure 2, intersect at the point $\beta_1 = .26667, \beta_2 = 3.1$ The value of $n$ for this position is $12$, and for $p$ is $1/6$ or $5/6$. These values of $\beta_1$ and $\beta_2$ are therefore for the binomials $(1/6 + 5/6)^{12}$, with the mean value $2$, and $(5/6 + 1/6)^{12}$ with the mean value $10$ .

From figure 2, it is seen that the hyperbolas extend into the area between the Poisson exponential line and the Type III line. Since, as stated above all $p$'s between $0$ and $1$ fall in the area between $\beta_1 = 0$ and the Poisson exponential line, it raises the question as to the meaning of the hyperbola outside that area. In the

point binomial $[p+(1-p)]^n$, there is nothing to force $p$ to lie between $0$ and $1$, and $n$ to be positive except the conditions we impose for applications to probability problems. If we consider the general case, without these limitations, an analysis of equation (3) shows that the radiating lines giving $\beta_1$ and $\beta_2$ for fixed values of $p$ continue into the area between the Poisson exponential and the Type III lines, the values of $p$ in this area being either negative, or the complements of these negative values, that is, positive values greater than $1$. This area includes all values of $p$ from $0$ to $-\infty$, and from $1$ to $+\infty$. Thus, lines giving $\beta_1$ and $\beta_2$ for all real values of $p$ from $-\infty$ to $+\infty$ are included between the vertical and the Type III lines. Outside this area, the values of $p$ are imaginary. Turning to the values of $n$, we can see from equation (6) that in the family of parallel lines giving $\beta_1$ and $\beta_2$ for fixed values of $n$ the lines below the Poisson exponential all have negative values of $n$. As $n$ approaches $-\infty$,

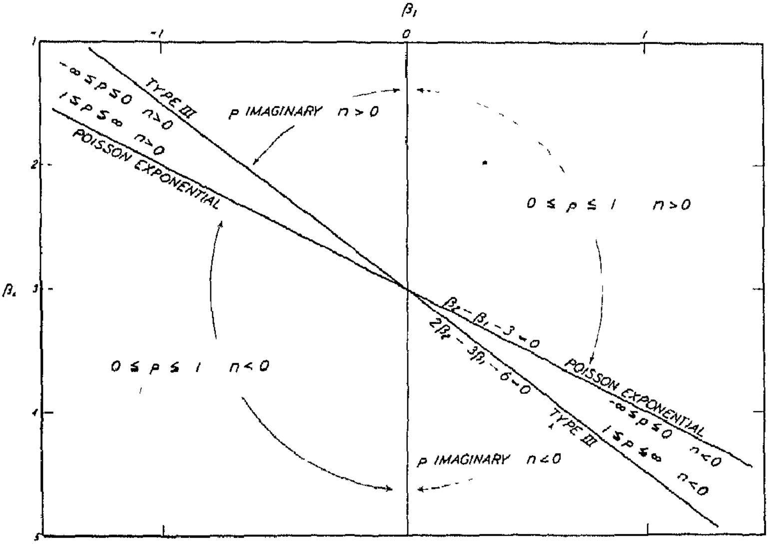



FIG 3 SUBDIVISION OF THE $\beta_1, \beta_2$ PLANE FOR BINOMIALS CLASSIFIED ACCORDING TO THE VALUES OF $p$ AND $n$

this line approaches the Poisson exponential line from below. Figure 3 shows the subdivisions of the $\beta_1 \beta_2$ plane for the various cases.

It follows from this discussion that the values of the hyperbola in the area between the Poisson exponential and Type III lines give $\beta$'s for point binomials with negative $n$'s, and since $np$ is a fixed positive value for any hyperbola, negative values of $p$, ranging from $0$ to $-\infty$. These results are in harmony with the case mentioned above, where the hyperbola degenerates into two straight lines (the Poisson exponential line and the Type III line) as the mean becomes indefinitely large, for $np$ approaches $\infty$ when either $n$ or $p$ approaches $\infty$ Thus, the Poisson exponential line is the limit when $n$ approaches $\infty$, and $p=1$, and the Type III line when $p$ approaches $-\infty$, and $n$ is negative.

5 *The relationship between $\beta_1$ and $\beta_2$ for constant values of $npq$.*

A further point of interest is the scatter of the $\beta$'s for point binomials of varying $p$'s but constant standard deviations. In equations (1) and (2), if we let $npq = \sigma^2$, and eliminate $p$ and $q$ we have the equation

$$2\beta_2 - 3\beta_1 - \frac{6\sigma^2 - 1}{\sigma^2} = 0 \tag{10}$$

This, like the lines giving $\beta_1$ and $\beta_2$ for constant values of $n$, is a family of parallel straight lines, but where the $n$ lines were parallel to the Poisson exponential line, this group is parallel to the Type III line. These lines intersect the radiating lines giving the $\beta$'s for constant values of $p$, and as $\sigma^2$ increases, the points of intersection of these lines approach the $\beta_1$ and $\beta_2$ for the normal curve. As $\sigma^2$ approaches $\infty$, the line given by (10) approaches the Type III line, and in the limiting case, crosses all the $p$ lines at the Gaussian position.

6. *Summary.*

Certain relationships between the third and fourth moments of the point binomial in terms of $\beta_1$ and $\beta_2$ have been discussed and the following results have been brought out:

A. For fixed values of $p$, $\beta_1$ and $\beta_2$ are linearly related, forming a family of radiating lines, all passing through the position of the $\beta$'s for the normal curve. Each of the lines represents the $\beta$'s for point binomials having a fixed value $p$ or its complement, $1-p$. The lines for values of $p$ between $0$ and $1$ are included between the vertical $\beta_1 = 0$, which is the line for $p = .5$ and the Poisson exponential line, $\beta_2 - \beta_1 - 3 = 0$, which is the line for $p = 0$ or $1$. The lines for negative values of $p$ or positive values greater than $1$, fall between the Poisson exponential line and the Type III line, $2\beta_2 - 3\beta_1 - 6 = 0$. For the rest of the plane; the values of $p$ are imaginary.

B. Although it has been shown by Pearson that in certain slope properties the skew point binomial resembles the Type III curve, none of the binomials which we interpret as probability functions, that is. those having $p$ between $0$ and $1$ and $n$ positive, has $\beta_1$ and $\beta_2$ approaching those of the Type III curve, except for the special case where the Type III curve becomes identical with the normal curve.

C. For fixed values of $n$, $\beta_1$ and $\beta_2$ determine a series of straight lines parallel to the Poisson exponential line. For positive values of $n$, these lines are above the Poisson exponential line, and for negative values below this line. Intersections of these lines with the radiating $p$ lines determine $\beta_1$ and $\beta_2$ for the point binomial of specified $p$ and $n$. As $n$ is held constant at increasingly great values, the points of intersection are closer and closer to the position of the $\beta$'s for the normal curve, and in the limiting case the line of $\beta$'s for constant $n$ intersects at the normal position all of the family of $p$ lines except the line for $p = 0$, with which it coincides.

D. $\beta_1$ and $\beta_2$ for point binomials of varying $p$ and $n$,

but constant mean values, lie on a family of hyperbolas, a particular hyperbola being determined by a specified mean. One of the asymptotes of all these hyperbolas is the Type III line, and the other is a line parallel to the Poisson exponential line, at a distance from it, depending on the value of the mean. The limit of this asymptote as the mean approaches $\infty$ is the Poisson exponential line. These hyperbolas are tangent to the line $\beta_1 = 0$, (the line for $p = .5$) and cross the other $p$ lines twice, the intersection of the hyperbola and any $p$ line or any $n$ line determining the $\beta$'s for the point binomial whose $p$ and $n$ are defined by the intersection.

E. For varying $p$ and $n$, but fixed $npq$, $\beta_1$ and $\beta_2$ lie on a family of straight lines parallel to the Type III line, one line in the group being determined by a particular $npq$. The limit of these lines as $npq$ is held constant at increasingly large values is the Type III line.

Margaret Merrell

# EDITORIAL

## NOTE ON THE COMPUTATION AND MODIFICATION OF MOMENTS

For the purpose of this note we shall deviate from the usual practice in the calculus of finite differences and define

$$\Delta u_x = u_{x+1} - M \cdot u_x ,$$

where $M$ is a constant. It follows that this generalized $\Delta$ and the symbol $E$ are connected by the operator relation

$$\Delta = (E - M), \text{ so that}$$

$$\Delta^n = (E - M)^n, \text{ and therefore}$$

$$(1) \quad \Delta^n u_x = u_{x+n} - \binom{n}{1} M \cdot u_{x+n-1} + \binom{n}{2} M^2 \cdot u_{x+n-2} - \binom{n}{3} M^3 \cdot u_{x+n-3} + \cdots .$$

If the $n$-$th$ unmodified moments about an arbitrary origin, and about the arithmetic mean, be designated by $\nu_n$ and $\bar{\nu}_n$, respectively, the usual relation may be written

$$(2) \quad \bar{\nu}_n = \nu_n - \binom{n}{1} M \cdot \nu_{n-1} + \binom{n}{2} M^2 \cdot \nu_{n-2} - \binom{n}{3} M^3 \cdot \nu_{n-3} + \cdots ,$$

where $M = \nu_1$ equals the distance of the mean from the provisional mean. From (1) and (2) it follows that

$$(3) \qquad \bar{\nu}_n = \Delta^n \nu_0 ,$$

that is, the $n$-$th$ moment about the mean is simply equal to the $n$-$th$ leading difference of $\nu_0$ . Since computing machines are ideally adapted to computations of the type $(A-BC)$, formula (3) is very effective.

As an illustration, let us compute the first seven moments about the mean for the distribution of weights of 7749 adult males born in the British Isles. (See Yule's "Theory of Statistics", p. 95.) The provisional mean is taken in the 150-lb. class, and the class interval of ten pounds is taken as the unit of $x$ .

TABLE 1

| $n$ | $\Sigma x^n f$ | $\nu_n$ | $\Delta\nu_n$ | $\Delta^2\nu_n$ | $\Delta^3\nu_n$ | $\Delta^4\nu_n$ |
|---|---|---|---|---|---|---|
| 0 | 7749 | 1 000000 | 000000 | 4.55356 | 6 92736 | 91.3249 |
| 1 | 1726 | 222738 | 4.55356 | 7.94161 | 92.8679 | 456 762 |
| 2 | 35670 | 4.60317 | 8 95586 | 94.6368 | 477.447 | 4471.10 |
| 3 | 77344 | 9 98116 | 96 6316 | 498.526 | 4577.45 | |
| 4 | 766026 | 98.8548 | 520.050 | 4688.49 | | |
| 5 | 4200496 | 542.069 | 4804.32 | | | |
| 6 | 38164290 | 4925.06 | | | | |

| $n$ | $\Delta^5\nu_n$ | $\Delta^6\nu_n$ |
|---|---|---|
| 1 | 436.420 | 4272.15 |
| 2 | 4369.36 | |

It is very important that the provisional mean be so chosen that $M = \nu_1$ is less than unity—otherwise particular attention must be paid to the number of digits which are significant in the values of the various differences.

Let us now discuss the modification of moments.

Designating the general modified or corrected moment by $\mu_n$ , Sheppard's formula for continuous variates may be written

$$(4) \qquad \mu_n = \nu_n - \binom{n}{2}\frac{1}{12}\nu_{n-2} + \binom{n}{4}\frac{7}{240}\nu_{n-4} - \binom{n}{6}\frac{31}{1344}\nu_{n-6} + \cdots,$$

so that for moments about the mean,

$$
(5)\qquad
\begin{cases}
\bar{\mu}_2 = \bar{\nu}_2 - \frac{1}{12} \\
\bar{\mu}_3 = \bar{\nu}_3 \\
\bar{\mu}_4 = \bar{\nu}_4 - \frac{1}{2}\bar{\nu}_2 + \frac{7}{240} \\
\bar{\mu}_5 = \bar{\nu}_5 - \frac{5}{6}\bar{\nu}_3 \\
\bar{\mu}_6 = \bar{\nu}_6 - \frac{5}{4}\bar{\nu}_4 + \frac{7}{16}\bar{\nu}_2 - \frac{31}{1344}
\end{cases}
$$

In many distributions discrete variates are grouped into classes, each class containing $k$ different values of the variable. The formula, corresponding to Sheppard's, for grouped-discrete distributions may be obtained by employing the calculus of finite differences, and was given without proof in an Editorial on page 111, Vol. 1, No. 1 of the Annals as follows,

$$
(6)\qquad
\begin{aligned}
\mu_n = \nu_n - \binom{n}{2}\frac{1-\frac{1}{k^2}}{12}\nu_{n-2} + \binom{n}{4}\frac{\left(1-\frac{1}{k^2}\right)\left(7-\frac{3}{k^2}\right)}{240}\nu_{n-4} \\
- \binom{n}{6}\frac{\left(1-\frac{1}{k^2}\right)\left(31-\frac{16}{k^2}+\frac{3}{k^4}\right)}{1344}\nu_{n-6} + \cdots\cdots.
\end{aligned}
$$

Obviously the limit of (6) as $k$ approaches infinity is (4). So

$$
(7)\qquad
\begin{cases}
\bar{\mu}_2 = \bar{\nu}_2 - \dfrac{1-\frac{1}{k^2}}{12} \\
\bar{\mu}_3 = \bar{\nu}_3 \\
\bar{\mu}_4 = \bar{\nu}_4 - \dfrac{1-\frac{1}{k^2}}{2}\bar{\nu}_2 + \dfrac{\left(1-\frac{1}{k^2}\right)\left(7-\frac{3}{k^2}\right)}{240} \\
\bar{\mu}_5 = \bar{\nu}_5 - \dfrac{5\left(1-\frac{1}{k^2}\right)}{6}\bar{\nu}_3 \\
\bar{\mu}_6 = \bar{\nu}_6 - \dfrac{5\left(1-\frac{1}{k^2}\right)}{4}\bar{\nu}_4 + \dfrac{\left(1-\frac{1}{k^2}\right)\left(7-\frac{3}{k^2}\right)}{16}\bar{\nu}_2 - \dfrac{\left(1-\frac{1}{k^2}\right)\left(31-\frac{16}{k^2}+\frac{3}{k^4}\right)}{1344}.
\end{cases}
$$

In both (5) and (7) above it is to be understood that the class interval, $\lambda$, is chosen as the unit of $x$.

The modified moments, $\bar{\mu}_n$, of the following table are obtained by applying formulae (5) to the values of $\bar{\nu}_n$, which were obtained as the leading differences of table 1.

TABLE 2

| $n$ | $\bar{\nu}_n$ | $\bar{\mu}_n$ | $\sigma^n$ | $\alpha_n = \frac{\bar{\mu}_n}{\sigma^n}$ |
|---|---|---|---|---|
| 0 | 1 00000 | 1.00000 | 1.00000 | 1.000000 |
| 1 | 00000 | .00000 | 2.114292 | .000000 |
| 2 | 4.55356 | 4.47023 | 4 47023 | 1 000000 |
| 3 | 6.92736 | 6 92736 | 9 45137 | .732948 |
| 4 | 91.3249 | 89.0773 | 19.9830 | 4 45765 |
| 5 | 436.420 | 430 647 | 42.2499 | 10.1929 |
| 6 | 4272 15 | 4159.96 | 89.3286 | 46.5692 |

Tables 3 and 4 shed an interesting light on the subject of modification, and are obtained from the results of formulae (5) and (7). If the class interval be denoted by $\lambda$, and the unmodified values of the standard deviation and skewness by $\sigma'$ and $\alpha_3'$, respectively, that is

$$(8)\qquad \begin{cases} \sigma_x' = \lambda\sqrt{\bar{\nu}_{2\cdot x}} \\ \alpha_3' = \dfrac{\bar{\nu}_{3\,x}}{(\sigma_x')^3} \end{cases}$$

it follows that

$$(9)\qquad \begin{cases} \sigma_x = \sigma_x'\,\omega \quad \text{and} \\ \alpha_3 = \dfrac{\alpha_3'}{\omega^3} \quad \text{where} \\ \omega = \sqrt{1 - \dfrac{1-\frac{1}{k^2}}{12}\left(\dfrac{\lambda}{\sigma_x'}\right)^2} \end{cases}$$

As mentioned before, the case of continuous variates is the special case for which $k = \infty$.

To illustrate' for our distribution taken from Yule, $\lambda = 10\ lbs.$, and by (8)

$$\sigma_v' = 10\sqrt{4\,55356\ lbs^2} = 21.\,3391\ lbs$$

$$\alpha_3' = \frac{6\,92736}{(2\,13391)^3} = .712919$$

$$\frac{\lambda}{\sigma_v'} = 468623$$

From table 3, $\frac{\lambda}{\sigma_v'} = 47,\ k = \infty$ we have $\omega = 9907$ and consequently $\sigma_v = 9907\ (21.3391\ lbs) = 21\,14\ lbs.$, agreeing with the more exact value deduced from table 2, i.e. $21\,14292\ lbs.$

From table 4, $\frac{\lambda}{\sigma'} = 47,\ k - \infty$ we have $\omega^{-3} = 1.028$ and therefore $\alpha_3 = 1.028(712919) = 7329$, again agreeing with the more accurate value of table 2.

By either interpolation of tables 3 and 4, or by direct computation of $\omega$ and $\omega^{-3}$, greater accuracy may be obtained

For an illustration of grouped-discrete variates we may refer to pages 32 and 37 of Vol. 1, No. 1 of the Annals For the so-called D(4.1) of Table IX, $\lambda = 4,\ \sigma_v' = 5\,089,\ \alpha_3' = 096$. Hence $\frac{\lambda}{\sigma_v'} = 79$, and since $k = 4$ the modified or adjusted values are

$$\sigma_v = 5\,089(9753) = 4\,963$$

$$\alpha_3 = .096(1.078) = 103.$$

It should be observed that the factors $\omega$ and $\omega^{-3}$ are independent of the number of variates, $N$, and are just as properly applied even if the frequency distribution method for computing the

standard deviation, skewness, etc is not employed. Thus, if I compute the standard deviation for the weights of ten individuals and those variates are recorded to the nearest pound, then the resulting

$$\sigma_v' = \sqrt{\frac{\Sigma(v-M_v)^2}{N}}$$

theoretically, if not practically, should be modified. If it developed that $\sigma_v' = 20\,lbs.$, then for weights to the nearest pound $\lambda = 1$, and

$$\omega = \sqrt{1-\frac{1}{12}\cdot\frac{1}{20^2}} = .999896$$

If, on the other hand, the weights had been taken to the nearest half pound, or to the nearest tenth pound, the corresponding values for $\omega$ would be 999974 and 999999 respectively Only if the variates be discrete and $k=1$, or if the variates be continuous and be measured with absolute accuracy (which is impossible from a practical point of view) so that $\lambda = 0$, can $\omega = 1$, and modification be ignored.

It should be clearly understood, however, that Sheppard's corrections are merely *expectations*. We have no assurance that the use of one of these corrections in any single instance will increase the accuracy of that determination—it is quite likely that in any isolated case the modification will introduce a still greater error into the calculation As pointed out in pages 36-38 of Vol. 1 of the Annals, and clearly revealed in the included table IX, modifying eliminates only the *systematic* errors, and ignores the *accidental* errors which may be numerically greater and of opposite algebraic sign than the correction itself.

Lastly, one must remember that the mathematical theory underlying Sheppard's Corrections assumes that both the frequency

function and a sufficient number of its derivatives vanish at the limits of the distribution. Consequently the case of J-shaped or U-shaped distributions, or of data not actually classified into such distributions, is not covered by formulae (5) or (7). In any event, it is evident that the practical necessity for modification in all cases depends upon the ratio of the limits of accuracy of the measurements *employed* in the computations to the unmodified standard deviation. If we throw accurately determined variates into frequency distributions with large class intervals, $\lambda$ ,—and thus simplify certain computations, then we must realize that the introduction of a systematic error is the penalty paid for such procedure, that the greater the value $\frac{\lambda}{\sigma'_v}$ , the greater the penalty, and that the practical necessity for modification rests entirely upon the accuracy demanded of the final results.

The chief value of tables 3 and 4 is that an inspection of these tables gives a rough idea of the value of modification so far as the standard deviation and the skewness are concerned.

Table 3

$$\omega = \sqrt{1 - \frac{1-\frac{1}{k^2}}{12}\left(\frac{\lambda}{\sigma'_\nu}\right)^2}$$

| $\frac{\lambda}{\sigma'_\nu}$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 01 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 02 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 03 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 04 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 |
| 05 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 |
| 06 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9999 | .9998 |
| 07 | .9998 | .9998 | .9998 | .9998 | .9998 | .9998 | .9998 | .9998 | .9998 | .9998 |
| 08 | .9998 | .9998 | .9998 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 |
| 09 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 |
| 10 | .9997 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 |
| 11 | .9996 | .9996 | .9995 | .9995 | .9995 | .9995 | .9995 | .9995 | .9995 | .9995 |
| 12 | .9995 | .9995 | .9994 | .9994 | .9994 | .9994 | .9994 | .9994 | .9994 | .9994 |
| 13 | .9995 | .9994 | .9993 | .9993 | .9993 | .9993 | .9993 | .9993 | .9993 | .9993 |
| 14 | .9994 | .9993 | .9992 | .9992 | .9992 | .9992 | .9992 | .9992 | .9992 | .9992 |
| 15 | .9993 | .9992 | .9991 | .9991 | .9991 | .9991 | .9991 | .9991 | .9991 | .9991 |
| 16 | .9992 | .9991 | .9990 | .9990 | .9990 | .9990 | .9989 | .9989 | .9989 | .9989 |
| 17 | .9991 | .9989 | .9989 | .9988 | .9988 | .9988 | .9988 | .9988 | .9988 | .9988 |
| 18 | .9990 | .9988 | .9987 | .9987 | .9987 | .9987 | .9987 | .9987 | .9987 | .9987 |
| 19 | .9989 | .9987 | .9986 | .9986 | .9985 | .9985 | .9985 | .9985 | .9985 | .9985 |
| 20 | .9987 | .9985 | .9984 | .9984 | .9984 | .9984 | .9984 | .9984 | .9983 | .9983 |
| 21 | .9986 | .9984 | .9983 | .9982 | .9982 | .9982 | .9982 | .9982 | .9982 | .9982 |
| 22 | .9985 | .9982 | .9981 | .9981 | .9980 | .9980 | .9980 | .9980 | .9980 | .9980 |
| 23 | .9983 | .9980 | .9979 | .9979 | .9979 | .9978 | .9978 | .9978 | .9978 | .9978 |
| 24 | .9982 | .9979 | .9977 | .9977 | .9977 | .9976 | .9976 | .9976 | .9976 | .9976 |
| 25 | .9980 | .9977 | .9976 | .9975 | .9975 | .9974 | .9974 | .9974 | .9974 | .9974 |
| 26 | .9979 | .9975 | .9974 | .9973 | .9973 | .9972 | .9972 | .9972 | .9972 | .9972 |
| 27 | .9977 | .9973 | .9971 | .9971 | .9970 | .9970 | .9970 | .9970 | .9970 | .9970 |
| 28 | .9975 | .9971 | .9969 | .9969 | .9968 | .9968 | .9968 | .9968 | .9968 | .9967 |
| 29 | .9974 | .9969 | .9967 | .9966 | .9966 | .9966 | .9965 | .9965 | .9965 | .9965 |
| 30 | .9972 | .9967 | .9965 | .9964 | .9963 | .9963 | .9963 | .9963 | .9963 | .9962 |
| 31 | .9970 | .9964 | .9962 | .9961 | .9961 | .9961 | .9961 | .9960 | .9960 | .9960 |
| 32 | .9968 | .9962 | .9960 | .9959 | .9958 | .9958 | .9958 | .9958 | .9958 | .9957 |
| 33 | .9966 | .9960 | .9957 | .9956 | .9956 | .9955 | .9955 | .9955 | .9955 | .9954 |
| 34 | .9964 | .9957 | .9955 | .9954 | .9953 | .9953 | .9952 | .9952 | .9952 | .9952 |
| 35 | .9962 | .9955 | .9952 | .9951 | .9950 | .9950 | .9950 | .9949 | .9949 | .9949 |
| 36 | .9959 | .9952 | .9949 | .9948 | .9947 | .9947 | .9947 | .9947 | .9946 | .9946 |
| 37 | .9957 | .9949 | .9946 | .9945 | .9944 | .9944 | .9944 | .9944 | .9943 | .9943 |
| 38 | .9955 | .9946 | .9943 | .9942 | .9941 | .9941 | .9941 | .9940 | .9940 | .9940 |
| 39 | .9952 | .9944 | .9940 | .9939 | .9938 | .9938 | .9937 | .9937 | .9937 | .9936 |
| 40 | .9950 | .9941 | .9937 | .9936 | .9935 | .9934 | .9934 | .9934 | .9934 | .9933 |
| 41 | .9947 | .9938 | .9934 | .9933 | .9932 | .9931 | .9931 | .9931 | .9930 | .9930 |
| 42 | .9945 | .9935 | .9931 | .9929 | .9928 | .9928 | .9927 | .9927 | .9927 | .9926 |
| 43 | .9942 | .9931 | .9928 | .9926 | .9925 | .9924 | .9924 | .9924 | .9923 | .9923 |
| 44 | .9939 | .9928 | .9924 | .9922 | .9921 | .9921 | .9920 | .9920 | .9920 | .9919 |
| 45 | .9937 | .9925 | .9921 | .9919 | .9918 | .9917 | .9917 | .9916 | .9916 | .9915 |
| 46 | .9934 | .9921 | .9917 | .9915 | .9914 | .9913 | .9913 | .9913 | .9912 | .9911 |
| 47 | .9931 | .9918 | .9913 | .9911 | .9910 | .9909 | .9909 | .9909 | .9908 | .9907 |
| 48 | .9928 | .9914 | .9910 | .9907 | .9906 | .9906 | .9905 | .9905 | .9905 | .9904 |
| 49 | .9925 | .9911 | .9906 | .9903 | .9902 | .9902 | .9901 | .9901 | .9901 | .9900 |
| 50 | .9922 | .9907 | .9902 | .9899 | .9898 | .9897 | .9897 | .9897 | .9896 | .9895 |

Table 3

$$\omega = \sqrt{1 - \frac{1-\frac{1}{k^2}}{12}\left(\frac{\lambda}{\sigma_v'}\right)^2}$$

| $\frac{\lambda}{\sigma_v'}$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|
| 51 | .9918 | .9903 | 9898 | 9895 | .9894 | 9893 | 9893 | .9892 | .9892 | 9891 |
| 52 | .9915 | 9899 | 9894 | 9891 | 9890 | 9889 | 9888 | .9888 | 9888 | .9887 |
| 53 | .9912 | 9895 | .9890 | 9887 | 9886 | 9885 | 9884 | .9884 | .9883 | .9882 |
| 54 | .9908 | .9891 | 9885 | 9883 | .9881 | 9880 | 9880 | .9879 | .9879 | .9878 |
| 55 | .9905 | 9887 | 9881 | 9878 | 9877 | 9876 | 9875 | 9875 | .9874 | .9873 |
| 56 | 9902 | 9883 | 9877 | 9874 | 9872 | .9871 | .9871 | 9870 | 9870 | 9868 |
| 57 | 9898 | 9879 | 9872 | 9869 | 9868 | 9867 | 9866 | 9865 | 9865 | .9864 |
| 58 | .9894 | 9875 | 9868 | 9865 | 9863 | .9862 | 9861 | 9861 | 9860 | 9859 |
| 59 | .9891 | 9870 | 9863 | 9860 | 9858 | 9857 | 9856 | 9856 | .9855 | .9854 |
| 60 | 9887 | 9866 | 9858 | .9855 | .9853 | .9852 | .9851 | .9851 | .9850 | 9849 |
| 61 | .9883 | .9861 | 9854 | 9850 | 9848 | 9847 | 9846 | .9846 | .9845 | .9844 |
| 62 | .9879 | .9857 | 9849 | 9845 | .9843 | 9842 | 9841 | 9841 | .9840 | .9839 |
| 63 | 9875 | 9852 | 9844 | .9840 | 9838 | 9837 | 9836 | 9835 | 9835 | .9833 |
| 64 | 9871 | 9847 | 9839 | 9835 | .9833 | .9832 | .9831 | .9830 | .9830 | .9828 |
| 65 | 9867 | .9842 | 9834 | 9830 | .9827 | 9826 | 9825 | .9825 | .9824 | 9822 |
| 66 | .9863 | 9837 | 9829 | .9824 | 9822 | .9821 | 9820 | .9819 | 9819 | 9817 |
| 67 | .9859 | 9832 | 9823 | .9819 | .9816 | 9815 | 9814 | 9814 | 9813 | 9811 |
| 68 | .9854 | .9827 | 9818 | .9813 | 9811 | .9808 | .9808 | 9808 | 9807 | .9805 |
| 69 | 9850 | .9822 | 9812 | 9803 | 9805 | 9804 | 9803 | 9802 | .9802 | 9800 |
| 70 | 9846 | 9817 | 9807 | 9802 | 9799 | 9798 | 9797 | 9796 | 9796 | 9794 |
| 71 | .9841 | .9811 | 9801 | 9796 | .9794 | 9792 | .9791 | 9790 | 9790 | .9788 |
| 72 | 9837 | 9806 | 9795 | .9790 | .9788 | 9786 | 9785 | .9784 | .9784 | .9782 |
| 73 | 9832 | .9801 | 9790 | 9785 | 9782 | 9780 | .9779 | 9778 | .9778 | .9775 |
| 74 | .9827 | 9795 | .9784 | 9779 | 9776 | 9774 | .9773 | 9772 | .9772 | .9769 |
| 75 | 9823 | 9789 | .9778 | .9772 | .9769 | 9768 | 9767 | .9766 | 9765 | .9763 |
| 76 | .9818 | 9784 | 9772 | .9766 | 9763 | .9761 | .9760 | .9759 | .9759 | .9756 |
| 77 | 9813 | 9778 | .9766 | 9760 | .9757 | 9755 | 9754 | .9753 | .9752 | .9750 |
| 78 | .9808 | .9772 | 9759 | .9754 | 9750 | 9749 | 9747 | .9746 | 9746 | .9743 |
| 79 | 9803 | .9766 | .9753 | .9747 | 9744 | .9742 | 9741 | .9740 | 9739 | .9736 |
| 80 | 9798 | 9760 | .9747 | .9741 | 9737 | .9735 | .9734 | .9733 | .9732 | 9730 |
| 81 | .9793 | 9754 | .9740 | 9734 | 9731 | 9729 | 9727 | 9726 | 9726 | .9723 |
| 82 | 9788 | .9748 | .9734 | .9727 | .9724 | 9722 | 9720 | .9719 | .9719 | 9716 |
| 83 | .9782 | 9742 | 9727 | 9721 | 9717 | 9715 | .9713 | 9712 | 9712 | .9709 |
| 84 | .9777 | .9735 | 9720 | 9714 | 9710 | 9708 | .9707 | .9706 | 9705 | .9702 |
| 85 | .9772 | .9729 | 9714 | 9707 | .9703 | 9701 | .9699 | 9698 | .9697 | .9694 |
| 86 | 9766 | .9722 | .9707 | .9700 | 9696 | 9694 | .9692 | 9691 | .9690 | .9687 |
| 87 | 9761 | .9716 | 9700 | .9693 | .9689 | .9687 | .9685 | .9684 | .9683 | .9680 |
| 88 | .9755 | 9709 | 9693 | .9685 | .9681 | .9679 | .9677 | .9676 | .9675 | .9672 |
| 89 | .9749 | .9702 | 9686 | .9678 | .9674 | .9672 | .9670 | .9669 | 9668 | .9664 |
| 90 | .9744 | .9695 | 9678 | 9671 | 9666 | 9664 | .9662 | .9661 | 9660 | 9657 |
| 91 | .9738 | .9688 | 9671 | 9663 | .9659 | 9656 | .9654 | 9653 | .9652 | .9649 |
| 92 | .9732 | 9681 | .9664 | .9656 | 9651 | .9649 | .9647 | .9646 | 9645 | .9641 |
| 93 | .9726 | .9674 | .9656 | .9648 | .9643 | 9641 | 9639 | 9638 | .9637 | 9633 |
| 94 | .9720 | .9667 | .9649 | 9640 | .9635 | .9633 | .9631 | 9630 | 9629 | .9625 |
| 95 | .9714 | .9660 | 9641 | .9632 | .9627 | .9625 | 9623 | 9622 | .9621 | 9617 |
| 96 | .9708 | .9653 | .9633 | 9624 | 9619 | 9617 | .9615 | 9613 | .9612 | .9608 |
| 97 | .9702 | 9645 | 9625 | .9616 | .9611 | .9608 | .9606 | .9605 | 9604 | .9600 |
| 98 | .9695 | .9638 | .9618 | .9608 | .9603 | 9600 | .9598 | 9597 | 9596 | .9591 |
| 99 | .9689 | .9630 | 9610 | 9600 | 9595 | .9592 | .9590 | .9588 | .9587 | .9583 |
| 100 | .9682 | .9623 | .9601 | .9592 | 9586 | 9583 | .9581 | .9580 | 9579 | .9574 |

Table 4

$$\omega^{-3} = \left[1 - \frac{1-\frac{1}{k^2}}{12}\left(\frac{\lambda}{\sigma'_v}\right)\right]^{-3}$$

| $\frac{\lambda}{\sigma'_v}$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 01 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 02 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 03 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 04 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 05 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 06 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.001 |
| 07 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 |
| 08 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 |
| 09 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 |
| 10 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 |
| 11 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.001 | 1.002 |
| 12 | 1.001 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 |
| 13 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 |
| 14 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 | 1.002 |
| 15 | 1.002 | 1.002 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 |
| 16 | 1.002 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 | 1.003 |
| 17 | 1.003 | 1.003 | 1.003 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 |
| 18 | 1.003 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 |
| 19 | 1.003 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 | 1.005 |
| 20 | 1.004 | 1.004 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 |
| 21 | 1.004 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 | 1.005 |
| 22 | 1.005 | 1.005 | 1.006 | 1.006 | 1.006 | 1.006 | 1.006 | 1.006 | 1.006 | 1.006 |
| 23 | 1.005 | 1.006 | 1.006 | 1.006 | 1.006 | 1.007 | 1.007 | 1.007 | 1.007 | 1.007 |
| 24 | 1.005 | 1.006 | 1.007 | 1.007 | 1.007 | 1.007 | 1.007 | 1.007 | 1.007 | 1.007 |
| 25 | 1.006 | 1.007 | 1.007 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 |
| 26 | 1.006 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 | 1.008 |
| 27 | 1.007 | 1.008 | 1.009 | 1.009 | 1.009 | 1.009 | 1.009 | 1.009 | 1.009 | 1.009 |
| 28 | 1.007 | 1.009 | 1.009 | 1.009 | 1.010 | 1.010 | 1.010 | 1.010 | 1.010 | 1.010 |
| 29 | 1.008 | 1.009 | 1.010 | 1.010 | 1.010 | 1.010 | 1.010 | 1.011 | 1.011 | 1.011 |
| 30 | 1.008 | 1.010 | 1.011 | 1.011 | 1.011 | 1.011 | 1.011 | 1.011 | 1.011 | 1.011 |
| 31 | 1.009 | 1.011 | 1.011 | 1.012 | 1.012 | 1.012 | 1.012 | 1.012 | 1.012 | 1.012 |
| 32 | 1.010 | 1.011 | 1.012 | 1.012 | 1.013 | 1.013 | 1.013 | 1.013 | 1.013 | 1.013 |
| 33 | 1.010 | 1.012 | 1.013 | 1.013 | 1.013 | 1.014 | 1.014 | 1.014 | 1.014 | 1.014 |
| 34 | 1.011 | 1.013 | 1.014 | 1.014 | 1.014 | 1.014 | 1.014 | 1.014 | 1.014 | 1.015 |
| 35 | 1.012 | 1.014 | 1.015 | 1.015 | 1.015 | 1.015 | 1.015 | 1.015 | 1.015 | 1.015 |
| 36 | 1.012 | 1.015 | 1.015 | 1.016 | 1.016 | 1.016 | 1.016 | 1.016 | 1.016 | 1.016 |
| 37 | 1.013 | 1.015 | 1.016 | 1.017 | 1.017 | 1.017 | 1.017 | 1.017 | 1.017 | 1.017 |
| 38 | 1.014 | 1.016 | 1.017 | 1.018 | 1.018 | 1.018 | 1.018 | 1.018 | 1.018 | 1.018 |
| 39 | 1.014 | 1.017 | 1.018 | 1.019 | 1.019 | 1.019 | 1.019 | 1.019 | 1.019 | 1.019 |
| 40 | 1.015 | 1.018 | 1.019 | 1.020 | 1.020 | 1.020 | 1.020 | 1.020 | 1.020 | 1.020 |
| 41 | 1.016 | 1.019 | 1.020 | 1.021 | 1.021 | 1.021 | 1.021 | 1.021 | 1.021 | 1.021 |
| 42 | 1.017 | 1.020 | 1.021 | 1.022 | 1.022 | 1.022 | 1.022 | 1.022 | 1.022 | 1.022 |
| 43 | 1.018 | 1.021 | 1.022 | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 | 1.024 |
| 44 | 1.018 | 1.022 | 1.023 | 1.024 | 1.024 | 1.024 | 1.024 | 1.024 | 1.024 | 1.025 |
| 45 | 1.019 | 1.023 | 1.024 | 1.025 | 1.025 | 1.025 | 1.025 | 1.026 | 1.026 | 1.026 |
| 46 | 1.020 | 1.024 | 1.025 | 1.026 | 1.026 | 1.027 | 1.027 | 1.027 | 1.027 | 1.027 |
| 47 | 1.021 | 1.025 | 1.026 | 1.027 | 1.027 | 1.028 | 1.028 | 1.028 | 1.028 | 1.028 |
| 48 | 1.022 | 1.026 | 1.028 | 1.028 | 1.029 | 1.029 | 1.029 | 1.029 | 1.029 | 1.029 |
| 49 | 1.023 | 1.027 | 1.029 | 1.030 | 1.030 | 1.030 | 1.030 | 1.030 | 1.030 | 1.031 |
| 50 | 1.024 | 1.028 | 1.030 | 1.031 | 1.031 | 1.031 | 1.032 | 1.032 | 1.032 | 1.032 |

Table 4

$$\omega^{-3} = \left[1 - \frac{1-\frac{1}{k^2}}{12} \cdot \left(\frac{\lambda}{\sigma_v'}\right)^2\right]^{-3}$$

| $\frac{\lambda}{\sigma_v'}$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 51 | 1.025 | 1.030 | 1.031 | 1.032 | 1.032 | 1.033 | 1.033 | 1.033 | 1.033 | 1.033 |
| 52 | 1.026 | 1.031 | 1.033 | 1.033 | 1.034 | 1.034 | 1.034 | 1.034 | 1.034 | 1.035 |
| 53 | 1.027 | 1.032 | 1.034 | 1.035 | 1.035 | 1.035 | 1.036 | 1.036 | 1.036 | 1.036 |
| 54 | 1.028 | 1.033 | 1.035 | 1.036 | 1.037 | 1.037 | 1.037 | 1.037 | 1.037 | 1.038 |
| 55 | 1.029 | 1.035 | 1.037 | 1.037 | 1.038 | 1.038 | 1.038 | 1.039 | 1.039 | 1.039 |
| 56 | 1.030 | 1.036 | 1.038 | 1.039 | 1.039 | 1.040 | 1.040 | 1.040 | 1.040 | 1.041 |
| 57 | 1.031 | 1.037 | 1.039 | 1.040 | 1.041 | 1.041 | 1.041 | 1.042 | 1.042 | 1.042 |
| 58 | 1.032 | 1.039 | 1.041 | 1.042 | 1.042 | 1.043 | 1.043 | 1.043 | 1.043 | 1.044 |
| 59 | 1.034 | 1.040 | 1.042 | 1.043 | 1.044 | 1.044 | 1.044 | 1.045 | 1.045 | 1.045 |
| 60 | 1.035 | 1.041 | 1.044 | 1.045 | 1.045 | 1.046 | 1.046 | 1.046 | 1.046 | 1.047 |
| 61 | 1.036 | 1.043 | 1.045 | 1.046 | 1.047 | 1.047 | 1.048 | 1.048 | 1.048 | 1.048 |
| 62 | 1.037 | 1.044 | 1.047 | 1.048 | 1.048 | 1.049 | 1.049 | 1.049 | 1.050 | 1.050 |
| 63 | 1.038 | 1.046 | 1.048 | 1.050 | 1.050 | 1.050 | 1.051 | 1.051 | 1.051 | 1.052 |
| 64 | 1.040 | 1.047 | 1.050 | 1.051 | 1.052 | 1.052 | 1.053 | 1.053 | 1.053 | 1.053 |
| 65 | 1.041 | 1.049 | 1.052 | 1.053 | 1.054 | 1.054 | 1.054 | 1.054 | 1.055 | 1.055 |
| 66 | 1.042 | 1.050 | 1.053 | 1.055 | 1.055 | 1.056 | 1.056 | 1.056 | 1.056 | 1.057 |
| 67 | 1.044 | 1.052 | 1.055 | 1.056 | 1.057 | 1.058 | 1.058 | 1.058 | 1.058 | 1.059 |
| 68 | 1.045 | 1.054 | 1.057 | 1.058 | 1.059 | 1.060 | 1.060 | 1.060 | 1.060 | 1.061 |
| 69 | 1.046 | 1.055 | 1.059 | 1.060 | 1.061 | 1.061 | 1.062 | 1.062 | 1.062 | 1.063 |
| 70 | 1.048 | 1.057 | 1.060 | 1.062 | 1.063 | 1.063 | 1.064 | 1.064 | 1.064 | 1.065 |
| 71 | 1.049 | 1.059 | 1.062 | 1.064 | 1.065 | 1.065 | 1.065 | 1.066 | 1.066 | 1.066 |
| 72 | 1.051 | 1.061 | 1.064 | 1.066 | 1.066 | 1.067 | 1.067 | 1.068 | 1.068 | 1.068 |
| 73 | 1.052 | 1.062 | 1.066 | 1.068 | 1.068 | 1.069 | 1.069 | 1.070 | 1.070 | 1.070 |
| 74 | 1.054 | 1.064 | 1.068 | 1.070 | 1.070 | 1.071 | 1.071 | 1.072 | 1.072 | 1.072 |
| 75 | 1.055 | 1.066 | 1.070 | 1.072 | 1.073 | 1.073 | 1.073 | 1.074 | 1.074 | 1.074 |
| 76 | 1.057 | 1.068 | 1.072 | 1.074 | 1.075 | 1.075 | 1.076 | 1.076 | 1.076 | 1.077 |
| 77 | 1.058 | 1.070 | 1.074 | 1.076 | 1.077 | 1.077 | 1.078 | 1.078 | 1.078 | 1.079 |
| 78 | 1.060 | 1.072 | 1.076 | 1.078 | 1.079 | 1.079 | 1.080 | 1.080 | 1.080 | 1.081 |
| 79 | 1.062 | 1.074 | 1.078 | 1.080 | 1.081 | 1.082 | 1.082 | 1.082 | 1.083 | 1.083 |
| 80 | 1.063 | 1.076 | 1.080 | 1.082 | 1.083 | 1.084 | 1.084 | 1.085 | 1.085 | 1.086 |
| 81 | 1.065 | 1.078 | 1.082 | 1.084 | 1.085 | 1.086 | 1.087 | 1.087 | 1.087 | 1.088 |
| 82 | 1.066 | 1.080 | 1.084 | 1.087 | 1.088 | 1.088 | 1.089 | 1.089 | 1.089 | 1.090 |
| 83 | 1.068 | 1.082 | 1.087 | 1.089 | 1.090 | 1.091 | 1.091 | 1.092 | 1.092 | 1.093 |
| 84 | 1.070 | 1.084 | 1.089 | 1.091 | 1.092 | 1.093 | 1.093 | 1.094 | 1.094 | 1.095 |
| 85 | 1.072 | 1.086 | 1.091 | 1.093 | 1.095 | 1.095 | 1.096 | 1.096 | 1.097 | 1.098 |
| 86 | 1.074 | 1.088 | 1.093 | 1.096 | 1.097 | 1.098 | 1.098 | 1.099 | 1.099 | 1.100 |
| 87 | 1.075 | 1.090 | 1.096 | 1.098 | 1.100 | 1.100 | 1.101 | 1.101 | 1.102 | 1.103 |
| 88 | 1.077 | 1.093 | 1.098 | 1.101 | 1.102 | 1.103 | 1.103 | 1.104 | 1.104 | 1.105 |
| 89 | 1.079 | 1.095 | 1.101 | 1.103 | 1.105 | 1.105 | 1.106 | 1.106 | 1.107 | 1.108 |
| 90 | 1.081 | 1.097 | 1.103 | 1.106 | 1.107 | 1.108 | 1.109 | 1.109 | 1.109 | 1.110 |
| 91 | 1.083 | 1.100 | 1.106 | 1.108 | 1.110 | 1.111 | 1.111 | 1.112 | 1.112 | 1.113 |
| 92 | 1.085 | 1.102 | 1.108 | 1.111 | 1.112 | 1.113 | 1.114 | 1.114 | 1.115 | 1.116 |
| 93 | 1.087 | 1.104 | 1.111 | 1.114 | 1.115 | 1.116 | 1.117 | 1.117 | 1.117 | 1.119 |
| 94 | 1.089 | 1.107 | 1.113 | 1.116 | 1.118 | 1.119 | 1.119 | 1.120 | 1.120 | 1.122 |
| 95 | 1.091 | 1.109 | 1.116 | 1.119 | 1.121 | 1.122 | 1.122 | 1.123 | 1.123 | 1.124 |
| 96 | 1.093 | 1.112 | 1.118 | 1.122 | 1.123 | 1.124 | 1.125 | 1.126 | 1.126 | 1.127 |
| 97 | 1.095 | 1.114 | 1.121 | 1.125 | 1.126 | 1.127 | 1.128 | 1.129 | 1.129 | 1.130 |
| 98 | 1.097 | 1.117 | 1.124 | 1.127 | 1.129 | 1.130 | 1.131 | 1.131 | 1.132 | 1.133 |
| 99 | 1.099 | 1.120 | 1.127 | 1.130 | 1.132 | 1.133 | 1.134 | 1.134 | 1.135 | 1.136 |
| 100 | 1.102 | 1.122 | 1.130 | 1.133 | 1.135 | 1.136 | 1.137 | 1.137 | 1.138 | 1.139 |

Vol. IV No. 4

# THE ANNALS
of
# MATHEMATICAL STATISTICS

*SIX DOLLARS PER ANNUM*

NOVEMBER 1933

PUBLISHED QUARTERLY BY

AMERICAN STATISTICAL ASSOCIATION

*Publication Office*—Edwards Brothers, Inc., Ann Arbor, Michigan

*Business Office*—530 Commerce Bldg., New York Univ., New York, N. Y.

*Entered as second class matter at the Postoffice at Ann Arbor, Mich., under the Act of March 3rd, 1879.*

STATEMENT OF OWNERSHIP

UNDER ACT OF CONGRESS OF AUGUST 24, 1912

*Publisher*—American Statistical Association, New York City, New York.

*Editor*—H. C. Carver, University of Michigan.

*Managing Editor*—Rattan S. Sekhon, Ann Arbor, Michigan.

*Business Manager*—J. W. Edwards, Ann Arbor, Michigan.

*Owner*—American Statistical Association, 530 Commerce Bldg, New York City.

# THE EXTENDED PROBABILITY THEORY FOR THE CONTINUOUS VARIABLE WITH PARTICULAR APPLICATION TO THE LINEAR DISTRIBUTION

*By*
H. P. LAWTHER, JR.

The engineering worker is often confronted with the necessity of utilizing a group of quantities concerning whose numerical values it is known only that they lie between definite upper and lower limits. If a number $n$ of specimens is selected from such a group and the sum of the $n$ values taken, intuition rules that there is negligible probability that this sum will be as great as $n$ times the upper limit or as small as $n$ times the lower limit, and that the most probable value must be intermediate between these two extremes. Some assurance is desired regarding the practical limits within which such a sum may be expected to fall. While the distribution of the individual values within their limits may be unknown directly, yet workable inferences frequently may be made from the nature of the quantities. For example, in many manufacturing operations it is economical to turn out items (such as bearing balls, paper condensers, or spacing washers) in large quantities with rather coarse precision. By means of gauges set to limits narrow as compared with the total spread, the product is then selected into bins, and in the operation of assembly a completed article utilizes the material from a single bin. The contents of any such bin clearly may be expected to follow a linear distribution very closely, and if the relative proportions of the product finding their ways into this bin and its immediate neighbors can be learned, the distribution may be specified with practical accuracy. The linear distribution is thus fundamental to a large class of problems.

On several occasions the writer's speculations have led to prob-

[illegible] the linear distribution, and the reference literature [illegible] for assistance. The special case of the rectangu- [illegible] seems first to have been formulated by Laplace[1] [illegible] century ago. Rietz,[2] Irwin,[3] Hall,[4] and Craig,[5] in [illegible] presented analyses applicable to the study of the [illegible] each from a somewhat different viewpoint. In [illegible] follow the logical processes of specialists in this [illegible] was driven to put forth considerable independent [illegible] to arrive at a satisfactory understanding. As the [illegible] effort still another angle of approach was developed. [illegible] employs steps and terminology familiar to one whose [illegible] education may have been limited to that commonly [illegible] a college engineering course, and should be readily [illegible] to a wide field of workers. Encouragement was [illegible] treating the case of the generalized linear distribu- [illegible] which appears to be new. In the application to [illegible] was necessary to carry out certain tedious compu- [illegible] interesting values and curves. The results of this [illegible] presented with the thought that they may stimulate the [illegible] and use of a law of considerable application to [illegible] practice.

[illegible] understood that when a selection, or the sum of n [illegible] spoken of there is meant the dimension of that selec- [illegible] of the dimensions of the n selections. Following

[1] [illegible] *Analytique des Probabilités,* Troisième Edition [illegible]

[2] [illegible] *Certain Law of Probability of Laplace,* Proceedings of [illegible] Mathematical Congress, Toronto (1924), vol. 2, pp. 795-

[3] [illegible] *Frequency Distribution of Means of Samples from a* [illegible] *Law of Frequency with Finite Moments, with special* [illegible] *Type II.* Biometrika, vol. 19 (1927), pp. 226-239.

[4] [illegible] *Distribution of Means of Samples of Size N drawn from a* [illegible] *variate takes values between 0 and 1, all such values* [illegible] *probable.* Biometrika, vol. 19 (1927), pp. 240-244.

[5] [illegible] *Distributions of Certain Statistics.* American Journal [illegible] vol. 54, No. 2 (1932), pp. 353-366

the usual notation, the symbol $f_n(x)$ will be defined to be such that the integral $\int_a^b f_n(x)\, dx$ is equal to the probability that the sum of $n$ selections lies between the values $a$ and $b$. Neglecting higher orders of infinitesimals, the probability that the sum of $n$ selections lies between $x$ and $x+\Delta x$ would then be equal to the product $f_n(x) \cdot \Delta x$ The sum of $n$ selections clearly is the sum of $n-1$ selections plus the value of an additional selection The probability that the sum of $n$ selections lies within the interval $x$ to $x + \Delta x$ must then be equal to the summation of the probabilities associated with all possible pairs of values for the sum of the first $n-1$ selections and for the last selection, respectively, that can yield a final sum lying between $x$ and $x+\Delta x$ The values $x - m\,\Delta x$ and $m\,\Delta x$, where $m$ is an integer, are such a pair, and the totality of these pairs is obtained by extending $m$ to all possible values. Recalling that the probability of the simultaneous occurrence of two independent events is equal to the product of the probabilities associated with their individual occurrences, there may be written in the conventional symbols

$$f_n(x) \cdot \Delta x = \sum_{m=-\infty}^{m=\infty} f_{n-1}(x - m\,\Delta x) \cdot \Delta x \; f_1(m\,\Delta x)\,\Delta x.$$

Setting $m\,\Delta x = \lambda$ , and passing to the limiting form, there is obtained

$$f_n(x) = \int_{-\infty}^{\infty} f_{n-1}(x-\lambda)\; f_1(\lambda)\; d\lambda,$$

as the general formula for determining all subsequent $f$'s from $f_1(x)$. The form of the function $f_1(x)$ is, of course, determined for any particular case from the best available physical data. The expression for $f_n(x)$ is then obtained by $n$-1 successive applications of the operation of integration indicated above. For the sake of

subsequent brevity, the operator $\mathrm{P}$ will be defined to be such that

$$\mathrm{P}\,\phi(x) = \int_{-\infty}^{\infty} \phi(x-\lambda)\cdot f_1(\lambda)\cdot d\lambda$$

Using this notation it would be written

$$f_n(x) = \mathrm{P}^{n-1}\cdot f_1(x).$$

For the general linear distribution $f_1(x)$ is given as follows:

$$f_1(x) = 0\ , \quad \text{for} \quad -\infty < x < 0$$

$$f_1(x) = 0\ , \quad \text{for} \quad a < x < \infty\ .$$

$f_1(x)$ is the equation of a straight line for $0 \leqq x \leqq a$, subject to the conditions: (1) the area under the line from $x=0$ to $x=a$ is unity, (2) no ordinate is negative for any value of $x$ in this interval. By imposing these conditions upon the general equation to a straight line there is obtained

$$f_1(x) = \frac{1}{a}\left[1+k-\frac{2K}{a}(a-x)\right]\ , \quad \text{for } 0 \leqq x \leqq a$$

where the parameter $K$ is restricted to the values $-1 \leqq K \leqq 1$. With $f_1(x)$ so defined it can be inferred immediately that $f_n(x)$ must be identically zero for all negative values of $x$, will have some positive finite value everywhere in the interval $0 < x < na$, and must be identically zero for all values of $x$ greater than $na$. Also, since $f_1(x)$ is discontinuous for $x \to 0$ and for $x \to a$ the application of the operator $\mathrm{P}$ must be effected through proper choice of limits of integration. In this connection three possible cases arise:

*Case 1:* Where $x$, the sum of $n$ selections, lies in the interval $0 \leqq x \leqq a$ it could have resulted only from a value for the sum of $n$-$1$ selections lying in the interval $x$ to $0$, coupled with a suitable value for the $n$-th selection lying in the interval $0$ to $x$. For this case the operator $\mathrm{P}$ will be distinguished as follows:

$$f_n(x) = \overset{x}{\underset{0}{\mathrm{P}}}\cdot f_{n-1}(x) = \int_0^x f_{n-1}(x-\lambda)\cdot f_1(\lambda)\cdot d\lambda\ , \quad \text{for } 0 \leqq x \leqq a.$$

*Case 2:* Where $x$, the sum of $n$ selections, lies in the interval $a \leqq x \leqq (n-1)a$ it could have resulted only from a value for the sum of $n$-1 selections lying in the interval $x$ to $x-a$, coupled with a suitable value for the $n$-th selection lying in the interval O to $a$. For this case the operator P will be distinguished as follows:

$$f_n(x) = \overset{a}{\underset{0}{\mathrm{P}}}\; f_{n-1}(x) = \int_0^a f_{n-1}(x-\lambda)\cdot f_1(\lambda)\cdot d\lambda\,, \text{ for } a \leqq x \leqq (n-1)a\,.$$

*Case 3:* Where $x$, the sum of $n$ selections, lies in the interval $(n-1)a \leqq x \leqq na$ it could have resulted only from a value for the sum of $n-1$ selections lying in the interval $(n-1)a$ to $x-a$, coupled with a suitable value for the $n$-th selection lying in the interval $x-(n-1)a$ to $a$. For this case the operator P will be distinguished as follows:

$$f_n(x) = \overset{a}{\underset{x\;(n-1)a}{\mathrm{P}}}\cdot f_{n-1}(x) = \int_{x-(n-1)a}^{a} f_{n-1}(x-\lambda)\; f_1(\lambda)\; d\lambda \text{ for, } (n-1)a \leqq x \leqq na\,.$$

The procedure now is analogous to that employed in establishing the binomial theorem The first few $f$'s are obtained by hand-power methods, until the sequences can be discerned and the expression for $f_n(x)$ can be inferred The expression for $f_n(x)$ is then established, first by applying to $f_n(x)$ the operator P and showing that this yields an expression for $f_{n+1}(x)$ wholly consistent with that for $f_n(x)$ when $n+1$ is substituted for $n$, and finally by showing that it degenerates into $f_1(x)$ when $n$ is taken as 1.

The preliminary steps, while very necessary, are quite tedious, and there would be no value in repeating them here. Suffice it to state that by such means it can be inferred that $f_n(x)$ is of the form

$$\begin{aligned} f_n(x) = \frac{1}{a^n}\cdot\left(-\frac{1}{p}\right)^{n-2} \Bigg\{ & \left(1+K+\frac{2K}{ap}\right)^{n}[na-x] - \binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)[na-a-x] \\ & + \binom{n}{2}\left(1+K+\frac{2K}{ap}\right)^{n-2}\left(1-K+\frac{2K}{ap}\right)^{2}[na-2a-x] \\ & - \binom{n}{3}\left(1+K+\frac{2K}{ap}\right)^{n-3}\left(1-K+\frac{2K}{ap}\right)^{3}[na-3a-x] \text{-----} \\ & \text{-----------------------------} \\ & + (-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n-1}[a-x] \Bigg\} \end{aligned}$$

where it is understood that each term including a bracket member of the form $[na-ba-x]$ is to be assigned the value zero for values of $x$ which render this negative. The use of brackets [* *] distinguishes the operand in each term. The symbol $\frac{1}{p}$ denotes the operation of integration with respect to $x$ between the upper limit $x$ and that lower limit for which the integrand vanishes. Thus $\frac{1}{p}\cdot[na-ba-x]^m = -\frac{1}{m+1}[na-ba-x]^{m+1}$. Where $\frac{1}{p}$ occurs with a negative exponent it signifies the inverse operation of differentiation with respect to $x$. The symbol $\binom{n}{b}$ means $\frac{n!}{b!(n-b)!}$ and is one of the familiar binomial coefficients.

Preparatory to establishing the validity of the inferred expression for $f_n(x)$, it is convenient to assemble certain working material. First $\varphi_n(x)$ will be defined as follows:

$$\begin{aligned}\varphi_n(x) = \frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-2}\Big\{&\left(1+K+\frac{2K}{ap}\right)^n[na-x] - \binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)[na-a-x]\\ &+\binom{n}{2}\left(1+K+\frac{2K}{ap}\right)^{n-2}\left(1-K+\frac{2K}{ap}\right)^2[na-2a-x]\\ &-\binom{n}{3}\left(1+K+\frac{2K}{ap}\right)^{n-3}\left(1-K+\frac{2K}{ap}\right)^3[na-3a-x] - - - -\\ &- - - - - - - - - - - - - - - -\\ &+(-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n-1}[a-x]\\ &+(-1)^n\left(1-K+\frac{2K}{ap}\right)^n[-x]\Big\}.\end{aligned}$$

where the symbols all have the same meaning as before, but here there is no special understanding regarding the bracket members of the form $[na-ba-x]$ and they are to exist for all values of $x$. Especial note should be made of the inclusion of a final term in $[-x]$. Otherwise the expression is identical in appearance with that for $f_n(x)$ for the interval $0 \leqq x \leqq a$. Next, the typical operation

$${}_0^a P\cdot[B-x] = \int_0^a [B-x+\lambda]\cdot\frac{1}{a}\cdot\left[1+K-\frac{2K}{a}(a-\lambda)\right]\cdot d\lambda\,,$$

is readily evaluated and yields

$$\frac{1}{a}\cdot\left(-\frac{1}{p}\right)\left\{\left(1+K+\frac{2K}{ap}\right)[B+a-x]-\left(1-K+\frac{2K}{ap}\right)[B-x]\right\}.$$

Now it can be written immediately that

$$\overset{a}{\underset{0}{P}}\cdot\varphi_n(x)=\frac{1}{a^{n+1}}\cdot\left(-\frac{1}{p}\right)^{n-1}\left\{\left(1+K+\frac{2K}{ap}\right)^{n+1}[na+a-x]-\left(1+K+\frac{2K}{ap}\right)^{n}\left(1-K+\frac{2K}{ap}\right)[na-x]\right.$$
$$-\binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n}\left(1-K+\frac{2K}{ap}\right)[na-x]$$
$$+\binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)^{2}[na-a-x]---$$
$$-----------------$$
$$---+(-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)^{2}\left(1-K+\frac{2K}{ap}\right)^{n-1}[2a-x]$$
$$-(-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n}[a-x]$$
$$+(-1)^{n}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n}[a-x]$$
$$\left.-(-1)^{n}\left(1-K+\frac{2K}{ap}\right)^{n+1}[-x]\right\}$$

Collecting terms, this becomes

$$\frac{1}{a^{n+1}}\cdot\left(-\frac{1}{p}\right)^{n+1-2}\left\{\left(1+K+\frac{2K}{ap}\right)^{n+1}[(n+1)a-x]\right.$$
$$-\binom{n+1}{1}\left(1+K+\frac{2K}{ap}\right)^{n+1-1}\left(1-K+\frac{2K}{ap}\right)[(n+1)a-a-x]$$
$$+\binom{n+1}{2}\left(1+K+\frac{2K}{ap}\right)^{n+1-2}\left(1-K+\frac{2K}{ap}\right)^{2}[(n+1)a-2a-x]---$$
$$-----------------$$
$$+(-1)^{n}\binom{n+1}{n}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n+1-1}[a-x]$$
$$\left.+(-1)^{n+1}\left(1-K+\frac{2K}{ap}\right)^{n+1}[-x]\right\},$$

and this is seen to correspond exactly with the expression for $\Phi_n(x)$ if $n+1$ is substituted for $n$. Consequently, $\varphi_n(x)$ must be the

result of $n-1$ successive applications of the operator ${}_{0}^{a}P$ to a certain $\phi_1(x)$ given by

$$\phi_1(x) = \frac{1}{a}\cdot\left(\frac{1}{p}\right)^{-1}\left\{\left(1+K+\frac{2K}{ap}\right)[a-x] - \left(1-K+\frac{2K}{ap}\right)[-x]\right\}$$

$$= \frac{1}{a}\cdot\left\{\left[1+K-\frac{2K}{a}(a-x)\right] - \left[1-K-\frac{2K}{a}(-x)\right]\right\} \equiv 0,$$

which is seen to be identically zero for all values of $x$. Now the application of the operator ${}_{0}^{a}P$ to zero yields zero. Therefore $\phi_n(x)$ must be identically zero for all values of $x$. Finally, it is convenient to evaluate the typical operation

$${}_{x-B}^{a}P\,[B-x] = \int_{x-B}^{a}[B-x+\lambda]\cdot\frac{1}{a}\left[1+K-\frac{2K}{a}(a-\lambda)\right]\cdot d\lambda\,.$$

This yields

$$\frac{1}{a}\left(-\frac{1}{p}\right)\left(1+K+\frac{2K}{ap}\right)[B+a-x]$$

The expression for $f_n(x)$ now may be established in straight-forward fashion.

In the interval $0 \leqq x \leqq a$ the expression for $f_n(x)$ will conclude with the term involving $[a-x]$. As has been shown before, the expression for $f_{n+1}(x)$ should then be given by

$$f_{n+1}(x) = {}_{0}^{x}P\cdot f_n(x), \quad \text{for} \quad 0 \leqq x \leqq a\,.$$

This operation may be evaluated readily, and will yield a result that is correct. The form of the result, however, is such that it does not display the desired correspondence with the expression for $f_n(x)$. The expression for $\phi_n(x)$ is introduced here to advantage Let it be written

$$f_{n+1}(x) = {}_{0}^{x}P\; f_n(x) + {}_{x}^{a}P\;\phi_n(x), \qquad \text{for } 0 \leqq x \leqq a\,.$$

Since the last term is identically zero its introduction is permissible. Remembering that $\phi_n(x)$ consists of all the terms of $f_n(x)$ plus a final term in $[-x]$, a rearrangement may be made giving

$$f_{n+1}(x) = {}^{a}_{0}P\ f_n(x) + {}^{a}_{0}P\ \frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-2}(-1)\left(1-K+\frac{2K}{ap}\right)^n[-x], \text{ for } 0 \leqq x \leqq a.$$

Using the operations that have been evaluated above, there is obtained immediately

$$\begin{aligned} f_{n+1}(x) = \frac{1}{a^{n+1}}\left(-\frac{1}{p}\right)^{n-1} \Bigg\{ &\left(1+K+\frac{2K}{ap}\right)^{n+1}[na+a-x] - \left(1+K+\frac{2K}{ap}\right)^n\left(1-K+\frac{2K}{ap}\right)[na-x] \\ &- \binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^n\left(1-K+\frac{2K}{ap}\right)[na-x] - \cdots \\ &+ \binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)^2[na-a-x] \\ &\cdots\cdots\cdots \\ \cdots &+ (-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)^2\left(1-K+\frac{2K}{ap}\right)^{n-1}[2a-x] \\ &- (-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^n[a-x] \Bigg\} \\ + \frac{1}{a^{n+1}}\left(-\frac{1}{p}\right)^{n-1} \Bigg\{ &(-1)^n\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^n[a-x] \Bigg\}, \text{for } 0 \leqq x \leqq a. \end{aligned}$$

Collecting terms, this becomes

$$\begin{aligned} f_{n+1}(x) = \frac{1}{a^{n+1}}\left(-\frac{1}{p}\right)^{n+1-2} \Bigg\{ &\left(1+K+\frac{2K}{ap}\right)^{n+1}[(n+1)a-x] \\ &- \binom{n+1}{1}\left(1+K+\frac{2K}{ap}\right)^{n+1-1}\left(1-K+\frac{2K}{ap}\right)[(n+1)a-a-x] \\ &+ \binom{n+1}{2}\left(1+K+\frac{2K}{ap}\right)^{n+1-2}\left(1-K+\frac{2K}{ap}\right)^2[(n+1)a-2a-x] - - \\ &\cdots\cdots\cdots \\ \cdots &+ (-1)^{n-1}\binom{n+1}{n-1}\left(1+K+\frac{2K}{ap}\right)^2\left(1-K+\frac{2K}{ap}\right)^{n-1}[2a-x] \\ &+ (-1)^n\binom{n+1}{n}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^n[a-x] \Bigg\}, \end{aligned}$$

for $0 \leqq x \leqq a$

and this is seen to be wholly consistent with the formula for $f_n(x)$ for the same interval with the substitution of $n+1$ for $n$.

Now for some interval between $x=a$ and $x=na$, say $(na-ba-a) \leqq x \leqq (na-ba)$ where $b$ is an integer having any value from zero to $n-2$, the expression for $f_n(x)$ will conclude with the term involving $[na-ba-x]$. For the interval immediately preceding, namely $(na-ba-2a) \leqq x \leqq (na-ba-a)$, the expression will include the additional term involving $[na-ba-a-x]$. Therefore in evaluating the operation ${}_0\overset{a}{P}\cdot f_n(x)$, which as has been shown before should yield the expression for $f_{n+1}(x)$, the integration of all but the $[na-ba-a-x]$ term will be carried over the complete range of $\lambda$ from $0$ to $a$. The term involving $[na-ba-a-x]$ will not enter into $f_n(x-\lambda)$ until $\lambda$ reaches the value $(x-na+ba+a)$, and so the integration of it will be between the limits $(x-na+ba+a)$ and $a$. Using the operations that have been evaluated previously there is obtained immediately

$$
\begin{aligned}
f_{n+1}(x) = \frac{1}{a^{n+1}}\left(-\frac{1}{p}\right)^{n-1} \Bigg\{ & \left(1+K+\frac{2K}{ap}\right)^{n+1}[na+a-x] - \left(1+K+\frac{2K}{ap}\right)^{n}\left(1-K+\frac{2K}{ap}\right)[na-x] \\
& - \binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n}\left(1-K+\frac{2K}{ap}\right)[na-x] \\
& + \binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)^{2}[na-a-x] - - - \\
& - - - - - - - - - - - - - - - - \\
& - - - + (-1)^{b}\binom{n}{b}\left(1+K+\frac{2K}{ap}\right)^{n-b+1}\left(1-K+\frac{2K}{ap}\right)^{b}[na-ba+a-x] \\
& - (-1)^{b}\binom{n}{b}\left(1+K+\frac{2K}{ap}\right)^{n-b}\left(1-K+\frac{2K}{ap}\right)^{b+1}[na-ba-x] \\
& + (-1)^{b+1}\binom{n}{b+1}\left(1+K+\frac{2K}{ap}\right)^{n-b}\left(1-K+\frac{2K}{ap}\right)^{b+1}[na-ba-x] \Bigg\},
\end{aligned}
$$

for $(na-ba-a) \leqq x \leqq (na-ba)$.

Collecting terms, this becomes

$$f_{n+1}(x)=\frac{1}{a^{n+1}}\cdot\left(-\frac{1}{p}\right)^{n+1-2}\left\{\left(1+K+\frac{2K}{ap}\right)^{n+1}[(n+1)a-x]\right.$$

$$-\binom{n+1}{1}\left(1+K+\frac{2K}{ap}\right)^{n+1-1}\left(1-K+\frac{2K}{ap}\right)[(n+1)a-a-x]--$$

$$\left.--+(-1)^{b+1}\binom{n+1}{b+1}\left(1+K+\frac{2K}{ap}\right)^{n+1-b-1}\left(1-K+\frac{2K}{ap}\right)^{b+1}[(n+1)a-(b+1)a-x]\right\},$$

$$\text{for } (na-ba-a)\leqq x\leqq(na-ba).$$

and this is seen to be wholly consistent with the expression for $f_n(x)$ for the same interval with the substitution of $n+1$ for $n$.

Finally, for the interval $na\leqq x\leqq(na+a)$ the expression for $f_n(x)$ is identically zero. For the interval immediately preceding, namely $(na-a)\leqq x\leqq na$, it consists of the single term involving $[na-x]$. Therefore

$$f_{n+1}(x)={}^{a}_{x-na}P\ f_n(x)={}^{a}_{x-na}P\cdot\frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-2}\left(1+K+\frac{2K}{ap}\right)^{n}[na-x]$$

$$=\frac{1}{a^{n+1}}\left(-\frac{1}{p}\right)^{n+1-2}\left(1+K+\frac{2K}{ap}\right)^{n+1}[(n+1)a-x]\ \text{for}\ na\leqq x\leqq(n+1)a$$

and this is seen to be wholly consistent with the expression for $f_n(x)$ for the corresponding interval.

Setting $n=1$ in the general expression for $f_n(x)$ for the interval $0\leqq x\leqq a$ there is obtained

$$f_1(x)=\frac{1}{a}\cdot\left(-\frac{1}{p}\right)^{-1}\left\{\left(1+K+\frac{2K}{ap}\right)[a-x]\right\}.$$

Carrying out the indicated operations, this becomes

$$f_1(x)=\frac{1}{a}\cdot\left[1+K-\frac{2K}{a}(a-x)\right],\ \text{for}\quad 0\leqq x\leqq a$$

and this is the $f_1(x)$ chosen at the start.

In the derivation of $f_{n+1}(x)$ from $f_n(x)$ it was assumed that the two forms of the operator $P$, namely ${}^{a}_{0}P$ and ${}^{a}_{x-B}P$, are commutative with the operator $\frac{1}{p}$ when applied to an operand of the

form $[B-x]$ It is very easy to show that ${}_0^a P \cdot \frac{1}{p}\,[B-x]$ yields a result identical with that of $\frac{1}{p}\,{}_0^a P\,[B-x]$ and that ${}_{x\,B}^{\;a} P \cdot \frac{1}{p}\,[B-x]$ yields a result identical with that of $\frac{1}{p}\,{}_{x\,B}^{\;a} P\,[B-x]$, and the space will not be taken here to give this demonstration The formula for $f_n(x)$ may thus be regarded as firmly established.

It has been shown that the complete expression $\phi_n(x)$ is identically zero for all values of $x$ Therefore, if in the interval $(na-ba-a) \leqq x \leqq (na-ba)$ the desired function $f_n(x)$ can be represented by the partial expression

$$f_n(x)=\frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-2}\left\{\left(1+K+\frac{2K}{ap}\right)^n[na-x]-\binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)[na-a-x]\right.$$
$$+\binom{n}{2}\left(1+K+\frac{2K}{ap}\right)^{n-2}\left(1-K+\frac{2K}{ap}\right)^2[na-2a-x]---$$
$$\left.----+(-1)^b\binom{n}{b}\left(1+K+\frac{2K}{ap}\right)^{n-b}\left(1-K+\frac{2K}{ap}\right)^b[na-ba-x]\right\},$$

it follows that it may equally well be represented in the same interval by the negative of the remainder of the complete expression, or

$$f_n(x)=\frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-2}\left\{-(-1)^{b+1}\binom{n}{b+1}\left(1+K+\frac{2K}{ap}\right)^{n-b-1}\left(1-K+\frac{2K}{ap}\right)^{b+1}[na-ba-a-x]\right.$$
$$-(-1)^{b+2}\binom{n}{b+2}\left(1+K+\frac{2K}{ap}\right)^{n-b-2}\left(1-K+\frac{2K}{ap}\right)^{b+2}[na-ba-2a-x]---$$
$$-----\;-(-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n-1}[a-x]$$
$$\left.-(-1)^n\left(1-K+\frac{2K}{ap}\right)^n[-x]\right\}.$$

Remembering that $\binom{n}{1}=\binom{n}{n-1}$, $\binom{n}{2}=\binom{n}{n-2}$, etc, this last may be rewritten

$$f_n(x)=\frac{1}{a^n}\left(\frac{1}{p}\right)^{n-2}\left\{\left(1-K+\frac{2K}{ap}\right)^n[x]-\binom{n}{1}\left(1-K+\frac{2K}{ap}\right)^{n-1}\left(1+K+\frac{2K}{ap}\right)[x-a]\right.$$
$$+\binom{n}{2}\left(1-K+\frac{2K}{ap}\right)^{n-2}\left(1+K+\frac{2K}{ap}\right)^2[x-2a]-----$$
$$\left.--+(-1)^{n-b-1}\binom{n}{n-b-1}\left(1-K+\frac{2K}{ap}\right)^{b+1}\left(1+K+\frac{2K}{ap}\right)^{n-b-1}[x-na+ba+a]\right\}.$$

where again it is understood that each term including a bracket member of the form $[x-ba]$ is to be assigned the value zero for values of $x$ which render this negative. The having of these two forms of expression for $f_n(x)$ is very valuable in computation work, since it limits the number of terms that have to be handled to $\frac{n+2}{2}$ at most.

Setting $K$ equal to zero gives the special case of the rectangular distribution, and the expressions for $f_n(x)$ reduce to the forms

$$f_n(x)=\frac{1}{a^n}\left(\frac{1}{p}\right)^{n-2}\left\{[na-x]-\binom{n}{1}[na-a-x]+\binom{n}{2}[na-2a-x]---\right\},$$

and

$$f_n(x)=\frac{1}{a^n}\cdot\left(\frac{1}{p}\right)^{n-2}\left\{[x]-\binom{n}{1}[x-a]+\binom{n}{2}[x-2a]-----\right\}.$$

Carrying out the indicated operations, these become

$$f_n(x)=\frac{1}{a^n(n-1)!}\left\{[na-x]^{n-1}-\binom{n}{1}[na-a-x]^{n-1}+\binom{n}{2}[na-2a-x]^{n-1}----\right\}$$

and

$$f_n(x)=\frac{1}{a^n(n-1)!}\left\{[x]^{n-1}-\binom{n}{1}[x-a]^{n-1}+\binom{n}{2}[x-2a]^{n-1}----\right\}.$$

This last expression is the one usually found in the literature, and it was originally developed by Laplace as the limiting form of an urn problem.

Setting $K$ equal to plus one or to minus one gives either of the extreme cases of the "right triangular distribution" For $K$ equal to plus one the expressions for $f_n(x)$ reduce to

$$f_n(x)=\frac{2^n}{a^n(n-1)!}\left\{\left(1+\frac{1}{ap}\right)^n[na-x]^{n-1}-\binom{n}{1}\left(1+\frac{1}{ap}\right)^{n-1}\left(\frac{1}{ap}\right)[na-a-x]^{n-1}\right.$$
$$\left.+\binom{n}{2}\left(1+\frac{1}{ap}\right)^{n-2}\left(\frac{1}{ap}\right)^2[na-2a-x]^{n-1}\quad\cdots\right\},$$

and

$$f_n(x)=\frac{2^n}{a^n(n-1)!}\cdot\left\{\left(\frac{1}{ap}\right)^n[x]^{n-1}-\binom{n}{1}\left(\frac{1}{ap}\right)^{n-1}\left(1+\frac{1}{ap}\right)[x-a]^{n-1}\right.$$
$$\left.+\binom{n}{2}\left(\frac{1}{ap}\right)^{n-2}\left(1+\frac{1}{ap}\right)^2[x-2a]^{n-1}---\right\}.$$

The function $f_n(x)$ normally has no direct practical application, but it is of interest to see its trend with increasing values of $n$. There are shown on Figure 1 several members of the family of curves originating with the rectangular distribution, and on Figure 2 corresponding members of the family originating with the right triangular distribution. In both figures the interval $a$ has been taken as unity, and there have actually been plotted the curves $y = n \cdot f_n(n \cdot x)$. This change in variable places all curves to a common base, and at the same time preserves the property of the total area under each being unity.

For the sum of $n$ selections the practical worker wants to know the minimum value $x'$; or the maximum value $x''$; or, most often of all, the shortest interval $x'$ to $x''$ associated with a certain probability value. The probability that the sum of $n$ selections will be less than $x'$ is given by

$$F_n(x') = \int_0^{x'} f_n(x) \cdot dx,$$

and the probability that the sum will exceed $x''$ is given by

$$F_n(x'') = \int_{x''}^{na} f_n(x)\ dx.$$

Noting that $\int_0^{x'} [x - ba]\ dx = \int_{ba}^{x'} [x-ba] \cdot dx = \frac{1}{p} \cdot [x' - ba]$ and that

$\int_{x''}^{na} [na - ba - x]\,dx = \int_{x''}^{na-ba} [na-ba-x]\,dx = -\frac{1}{p} \cdot [na - ba - x'']$, since the bracket members are assigned the value zero for values of $x$ which render them negative, there may be written immediately

$$F_n(x') = \frac{1}{a^n} \cdot \left(\frac{1}{p}\right)^{n-1} \left\{ \left(1 - K + \frac{2K}{ap}\right)^n [x'] - \binom{n}{1} \left(1 - K + \frac{2K}{ap}\right)^{n-1} \left(1 + K + \frac{2K}{ap}\right) [x' - a] \right.$$
$$\left. + \binom{n}{2} \left(1 - K + \frac{2K}{ap}\right)^{n-2} \left(1 + K + \frac{2K}{ap}\right)^2 [x' - 2a] --- \right\},$$

and

$$F_n(x'') = \frac{1}{a^n} \cdot \left(-\frac{1}{p}\right)^{n-1} \left\{ \left(1 + K + \frac{2K}{ap}\right)^n [na - x''] \right.$$
$$\left. - \binom{n}{1} \left(1 + K + \frac{2K}{ap}\right)^{n-1} \left(1 - K + \frac{2K}{ap}\right) [na - a - x''] + \binom{n}{2} \left(1 + K + \frac{2K}{ap}\right)^{n-2} \left(1 - K + \frac{2K}{ap}\right)^2 [na - 2a - x''] -- \right\}.$$

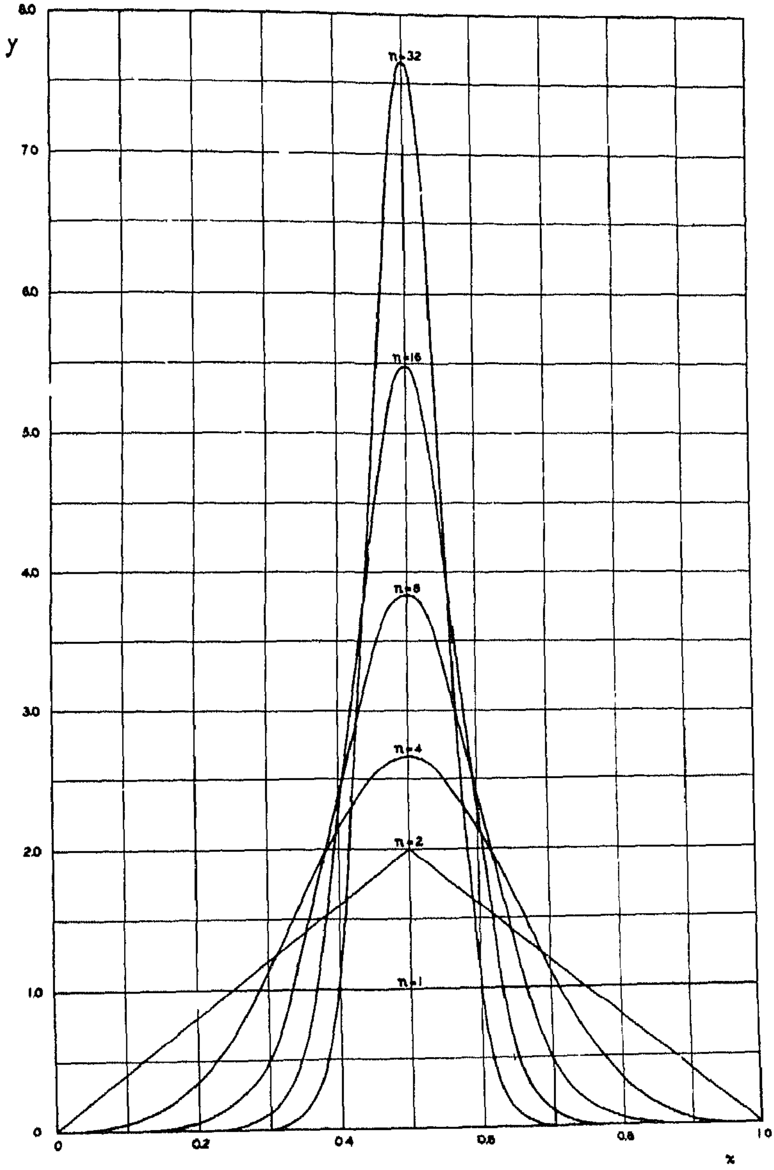



FIGURE -# 1

GRAPHS OF CURVES $y = n\, f_n(n\, x)$ FOR RECTANGULAR DISTRIBUTION

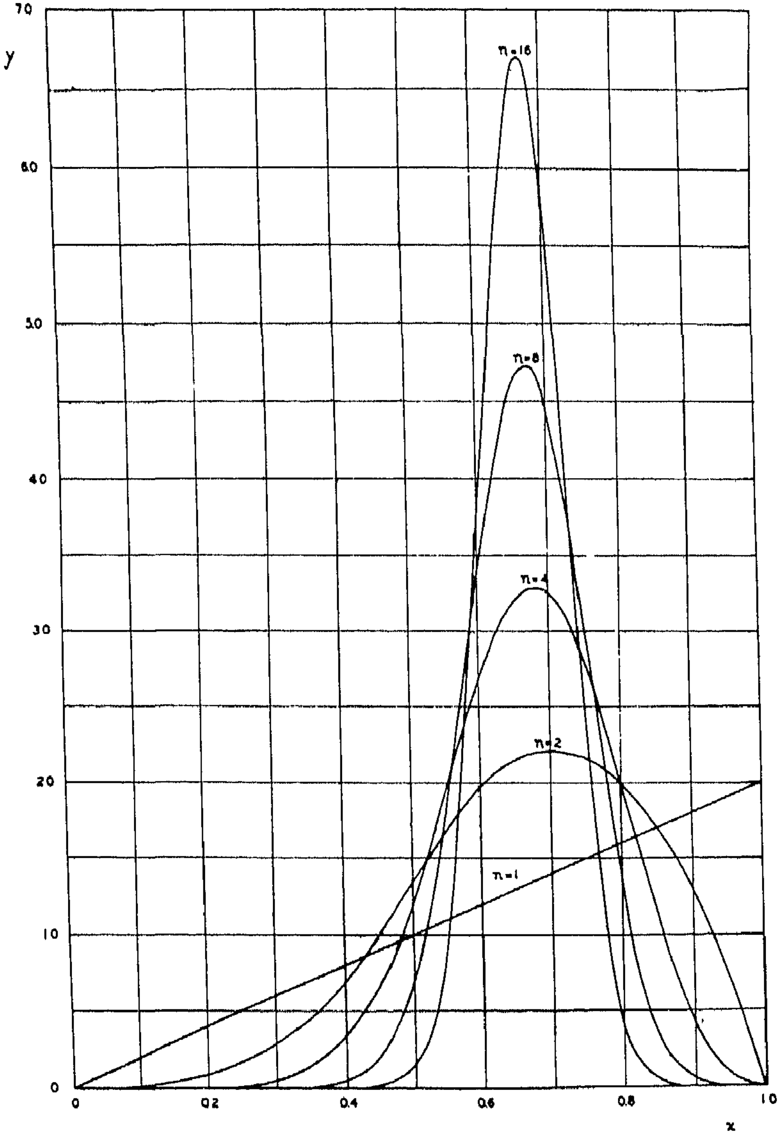



FIGURE-#2

GRAPHS OF CURVES $y = n\,f_n(nx)$ FOR RIGHT TRIANGULAR DISTRIBUTION

For $x' = na$ or for $x'' = 0$ these last expressions respectively should equate to unity, the total area under any probability curve. It is simple to verify that they do so, and the demonstration will be given for one of them. For the interval $0 \leqq x'' \leqq a$ the expression for $F_n(x'')$ concludes with the term containing $[a-x'']$. Let there be added and subtracted a term involving $[-x'']$ to give

$$F_n(x'') = \frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-1}\left\{\left(1+K+\frac{2K}{ap}\right)^n[na-x'']\right.$$
$$-\binom{n}{1}\left(1+K+\frac{2K}{ap}\right)^{n-1}\left(1-K+\frac{2K}{ap}\right)[na-a-x''] - \cdots + (-1)^{n-1}\binom{n}{n-1}\left(1+K+\frac{2K}{ap}\right)\left(1-K+\frac{2K}{ap}\right)^{n-1}[a-x'']$$
$$\left. + (-1)^n\left(1-K+\frac{2K}{ap}\right)^n[-x'']\right\} - \frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-1}(-1)^n\left(1-K+\frac{2K}{ap}\right)^n[-x''].$$

From inspection it is seen that this may be written

$$F_n(x'') = {}_0^a P^{n-1}\cdot\left(-\frac{1}{p}\right)\ \Phi_1(x'') - \frac{1}{a^n}\left(-\frac{1}{p}\right)^{n-1}(-1)^n\left(1-K+\frac{2K}{ap}\right)^n[-x''],$$

where $\Phi_1(x'')$ is the expression introduced previously. Upon carrying out the operations it is found that $-\frac{1}{p}\ \Phi_1(x'') = 1$, and ${}_0^a P\ [1] = 1$. It follows therefore that

$$F_n(x'') = 1 - \frac{1}{a^n}\cdot\left(-\frac{1}{p}\right)^{n-1}(-1)^n\left(1-K+\frac{2K}{ap}\right)^n[-x'']\ .\text{for}\quad 0 \leqq x'' \leqq a$$

and it is apparent now that for $x'' = 0$ this expression is equal to unity. Thus it is seen that the function $F_n(x)$ also possesses an end-for-end symmetry similar to that of $f_n(x)$. The complete expression corresponding to $F_n(x)$ is equal to unity instead of zero, however, and where the desired function is represented by a partial expression it can also be equally well represented by one minus the remaining terms of the complete expression.

Referring to Figure 3 it is seen that the sum of area A plus area B, or $F_n(x') + F_n(x'')$, gives the probability that

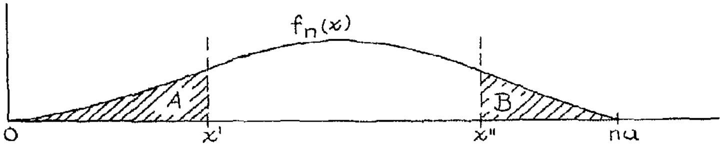



Figure 3

the sum of $n$ selections lies outside the interval $x'$ to $x''$. In an actual problem usually either the sum of A and B would be assigned and it would be desired to find the interval $x'$ to $x''$ associated with this expectancy, or the interval $x'$ to $x''$ would be assigned and it would be desired to find the expectancy associated with these limits. With $f_n(x)$ of the character shown in Figure 3 and with the sum alone of A and B fixed, any number of pairs of values are possible for $x'$ and $x''$. It is also clear that the length of the interval $x''-x'$ will depend upon the relative magnitudes of A and B. There are two special cases, however, which cover all normal demands. It is seen at a glance that the interval $x''-x'$ will be shortest for a given A plus B when $f_n(x'')=f_n(x')$. Purely from the standpoint of deviations, this shortest interval represents the optimum results of which the group is capable. Where the absolute magnitude is of primary concern it might be specified that $x'=na-x''$.

The function which is of final interest, then, is represented by the sum $F_n(x')+F_n(x'')$, subject either to the restriction that $f_n(x')=f_n(x'')$ or to the restriction that $x'=na-x''$. For the special case of the rectangular distribution $f_n(x)$ is symmetrical about the line $x=\frac{na}{2}$, and the two restrictions imply the same thing. The symmetry of the rectangular distribution permits giving formal expression to the sum $F_n(x')+F_n(x'')$ as a function of $(x''-x')$ when $x'=na-x''$. Under this condition the sum becomes equal to $2F_n(x')$ and $x'$ may be written as $\left[\frac{na-(x''-x')}{2}\right]$ and there is obtained

$$F_n(x')+F_n(x'')=2F_n\left[\frac{na-(x''-x'')}{2}\right], \quad \text{for } x'=na-x''$$

For linear distributions other than the rectangular this simplicity of expression is not possible.

On Figures 4 and 5 are shown graphs of the curves $y = F_n(n\,x')$ for several values of $n$ for the rectangular and for the right triangular distribution respectively. Here again the interval $a$ has been taken as unity and change in variable has been made to place all curves to a common base. The values of $F_n(n\,x'')$ may be read directly from the same curves, since $F_n(n \cdot x'') = 1 - F_n(n\,x')$ for corresponding values of $x'$ and $x''$. Finally, on Figures 6 and 7 are shown curves for the sum $\left[F_n(x') + F_n(x'')\right]$ plotted as a function of $\left(\frac{x''-x'}{na}\right)$, subject to the restriction that $f_n(x') = f_n(x'')$, for several values of $n$ for the rectangular and for the right triangular distribution respectively. The values for Figure 6 were computed directly from the formula given in the paragraph above. For Figure 7, however, the values were derived graphically from Figures 2 and 5.

Figures 6 and 7 are applicable immediately to practical problems. As a simple example, suppose there is at hand a group whose individuals are known to lie within the limits of $D$ and $D+a$ and to follow a right triangular distribution with the larger probability associated with the larger limit, and it is desired to know for the sum of eight selections what limits may be expected to be associated with a probability value of 0.01. Referring to Figure 7 it is seen that the curve for $n = 8$ reaches an ordinate value of 0.01 at an abscissa value of approximately 0.45. Referring now to Figure 2, the distance 0.45 is fitted in between the two legs of the curve for $n = 8$, and values for $\frac{x'}{8}$ and $\frac{x''}{8}$ of 0.43 and 0.88 are found. Consequently it can be concluded that for sums of eight selections from this group the probability is 0 01 that the values will lie outside the interval $8D + 3.44a$ to $8D + 7 04a$.

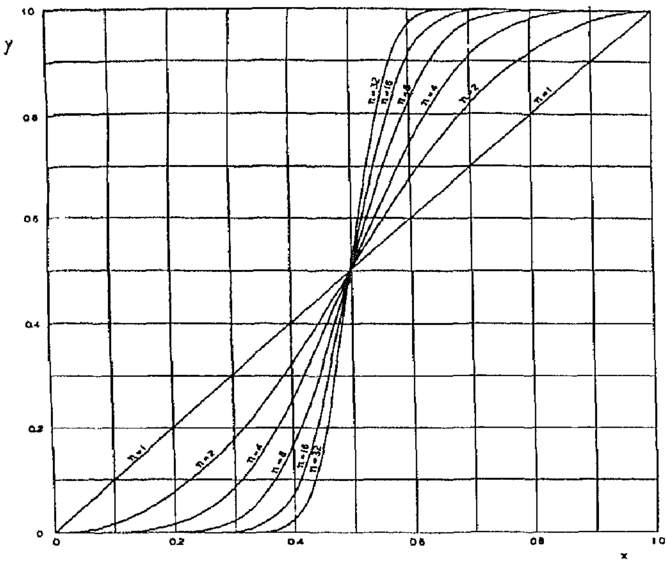



FIGURE-8.4

GRAPHS OF CURVES $y = F_n(n, x)$ FOR RECTANGULAR DISTRIBUTION

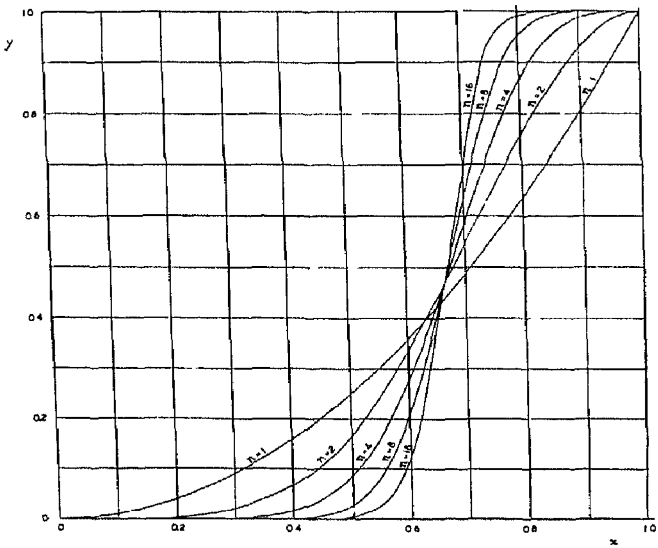



FIGURE-8.5

GRAPHS OF CURVES $y = F_n(n, x)$ FOR RIGHT TRIANGULAR DISTRIBUTION

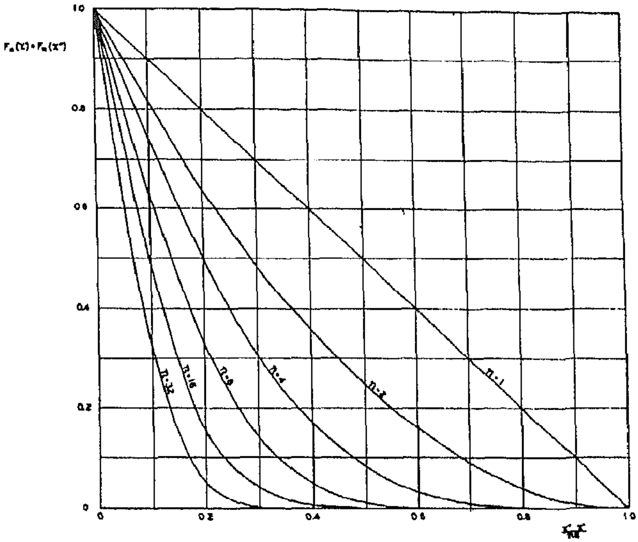



FIGURE-#6

GRAPHS OF CURVES OF $\left[F_n(x) + F_n(x'')\right]$ VS. $\left[\frac{x'-x}{n\epsilon}\right]$ FOR RECTANGULAR DISTRIBUTION FOR CONDITION THAT $f_n(x) = f_n(x'')$

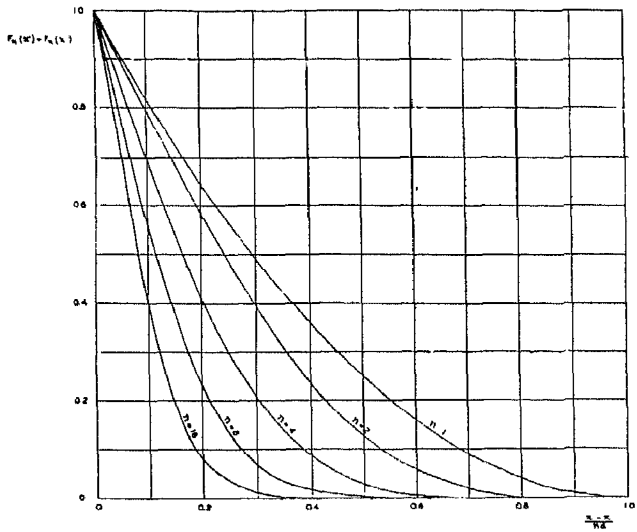



FIGURE-#7

GRAPHS OF CURVES OF $\left[F_n(x) + F_n(x')\right]$ VS $\left[\frac{x-x}{n\epsilon}\right]$ FOR RIGHT TRIANGULAR DISTRIBUTION FOR CONDITION THAT $f_n(x') = f_n(x'')$

While this study has been concerned primarily with the linear distribution, it is obvious that the results may find occasional wider application. The curves for $f_2(x)$, $f_4(x)$, etc., are of a character suggestive of distributions that might occur not infrequently in engineering and physics. If any one of them, say $f_n(x)$, is found to fit the group at hand with practical accuracy, then the sequence $f_n(x)$, $f_{2n}(x)$, $f_{3n}(x)$, etc. clearly will give the distribution curves associated with the sums of one, two, three, etc., selections from this group.

H. P. Lawther, Jr

# ON THE ELIMINATION OF SYSTEMATIC ERRORS DUE TO GROUPING

*By*
JOHN R. ABERNETHY

In the calculation of the moments of a frequency distribution it is often desirable, or even necessary, to consider not the distribution itself but another derived from it by certain groupings. As a first approximation to the moments of the original distribution we take the corresponding moment of the grouped distribution. But this first approximation is not satisfactory, and it is necessary to obtain some method for the elimination of part of the error committed in replacing the moments of the original distribution by the corresponding moments of the grouped distribution.

This problem was first discussed by W. F. Sheppard in a paper: *On the Calculation of the most Probable Value of Frequency-Constants, for Data arranged according to Equidistant Divisions of a Scale.*[1] If we denote the n-th moment of the original distribution by $\mu_n$ and the n-th moment of the grouped distribution by $\nu_n$, we will have Sheppard's corrections in the form:

$$\mu_1 = \nu_1 = 0,$$

$$\mu_2 = \nu_2 - \frac{1}{12},$$

$$\mu_3 = \nu_3,$$

$$\mu_4 = \nu_4 - \frac{1}{2}\nu_2 + \frac{7}{240},$$

$$\mu_5 = \nu_5 - \frac{5}{6}\nu_3, \qquad \text{etc.}$$

As pointed out by Karl Pearson[2] the hypotheses under which these formulae have been obtained are: (a) that Taylor's theorem

[1] Proceedings London Mathematical Society, Vol. 29, p. 353-380.

[2] *On an elementary proof of Sheppard's formulae for correcting raw moments and other allied points*, editorial in Biometrica, Vol. 3, p. 308-312.

may be applied to the frequency function throughout the range; (b) $x^{i}f^{(i)}(x)$ is finite and continuous throughout the range, (c) $f(x)$ and its derivatives vanish at the limits of the range These hypotheses are not always satisfied by the frequency functions with which the statistician has to work, and as it is impossible to tell before calculating the moments of a distribution whether the corresponding theoretical frequency function satisfies these conditions, it is desirable to study the problem from another standpoint

A comparison of the title of Sheppard's paper and the paper itself suggests the question, in what sense do Sheppard's formulae give the most probable value of the moments of a distribution? A partial answer is given by B. L. Shook in the *Synopsis of Elementary Mathematical Statistics.*[3] Miss Shook presents[4] the formulae $\mu_1 = \nu_1 = 0$, $\mu_2 = \nu_2 - \frac{1}{12}\left(1 - \frac{1}{m^2}\right)$, and $\mu_3 = \nu_3$ for a discrete distribution with $m$ values of the variable grouped in each class interval and shows that for a particular distribution these formulae serve to eliminate the systematic errors from $M$, $\mu_2$, and $\mu_3$. Two problems are suggested by the *synopsis*: the derivation of formulae for the class of discrete distributions, as these three formulae are stated without proof;[5] the proof that this larger set of formulae and those of Sheppard do serve under all conditions to eliminate the systematic error due to grouping, subject only to the existence of the moments involved. When we have solved these two problems, we shall be in position to understand the true nature of the approxi-

[3] THE ANNALS OF MATHEMATICAL STATISTICS, Vol 1, p 34-40

[4] These formulae are only special cases of a more general formula stated by H. C. Carver in an editorial ANNALS OF MATHEMATICAL STATISTICS, Vol. 1; formula (14), p. 111

[5] Two methods of developing this formula suggest themselves: (a) the elimination of the moment of a continuous graduating function expressed in terms of fine groupings of class intervals of $\frac{1}{m}$ on the one hand and in the terms of coarse groupings of unit class intervals on the other, (b) by a process similar to that of Sheppard, employing for example Lubbock's formula instead of the Euler-Maclaurin sum formula. According to a statement made by Professor Carver, the formulae in question were derived by the latter process.

mation involved in employing Sheppard's corrections and corrections similar to them for discrete distributions.

The problem we wish to consider is this Given the probabilities $\phi(x_i)$ that a value of the statistical variable $x$ taken at random will fall within the interval $x_i - \frac{1}{2} < x < x_i + \frac{1}{2}$, we wish to find the moments of the distribution We consider this problem for two classes of distributions: the distribution of a discrete variable; the distribution of a continuous variable. In either case we shall work with the uni-frequency distributional function $f(x)$. For the discrete distribution $\frac{1}{m} f\left(\frac{j}{m}\right)$ represents the probability that a value of $x$ taken at random will be the number $\frac{j}{m}$; $m$ denotes a definite positive integer, $(j = -2, -1, 0, 1, 2, \ldots)$. For the continuous distribution $\int_a^b f(x)dx$ represents the probability that a value of $x$ taken at random will fall within the interval $a < x < b$ Thus the function $f(x)$ has the value zero outside the range of the distribution and we may for the sake of convenience denote the limits of summation and integration as $\pm\infty$. For the $n$-th moment about the origin we have.

$$\mu_n' = \sum_{j=-\infty}^{+\infty} \left(\frac{j}{m}\right)^n f\left(\frac{j}{m}\right) \cdot \frac{1}{m},$$

for the discrete distribution, and

$$\mu_n' = \int_{-\infty}^{+\infty} x^n f(x)\, dx,$$

for the continuous distribution. What we want is the value of $\mu_n'$. What we are able to find is the value of

$$\nu_n' = \sum_{i=-\infty}^{+\infty} (x_i)^n \phi(x_i).$$

In establishing approximate relations between the set of true moments $\{\mu_n'\}$ and the set of raw moments $\{\nu_n'\}$ we shall employ another set of statistical constants $\{\bar{\nu}_n'\}$. For the discrete distribution there are $m$ distinct sets of groupings that can be made, leading to $m$ values of the raw moment $\nu_n'$; $\bar{\nu}_n'$ is used to represent the average of these. Similarly for the continuous distribution, $\bar{\nu}_n'$ is used to denote the average of the moments $\nu_n'$ corresponding to $x_i = i+t$ for all values of $t$ satisfying $0 \leq t < 1$. We shall call this intermediate set of statistical constants $\{\bar{\nu}_n'\}$ the average grouped moments of the distribution. We then divide the problem into two parts. First we seek the expression of $\mu_n'$ in terms of the $\{\bar{\nu}_n'\}$. Secondly we seek the nature of approximation in replacing $\bar{\nu}_n'$ by $\nu_n'$. The first of these can be solved completely without approximation and without any assumption other than the existence of the moments involved. We can best understand the nature of the approximation involved in the second after the first of our two problems has been solved.

The $m$ values of $\nu_n'$ corresponding to the $m$ distinct methods of grouping a discrete distribution are given by

$$\nu_n'(t) = \sum_{i=-\infty}^{+\infty} \left(i+t+\frac{m-1}{2m}\right)^n \sum_{j=0}^{m-1} \frac{1}{m} f\left(i+t+\frac{j}{m}\right);\ mt = 0, 1, \ldots\ m-1$$

The average of these is

$$\bar{\nu}_n' = \frac{1}{m} \sum_{k=0}^{m-1} \sum_{i=-\infty}^{+\infty} \left(i+\frac{k}{m}+\frac{m-1}{2m}\right)^n \sum_{j=0}^{m-1} \frac{1}{m} f\left(i+\frac{k+j}{m}\right).$$

We shall first express the average grouped moment $\bar{\nu}_n'$ in terms of the true moments $\{\mu_n'\}$, and then solve for the $\{\mu_n'\}$ in terms of the $\{\bar{\nu}_n'\}$. We wish to arrange the right hand side of the above equation according to values of the argument $x$ appearing in $f(x)$; we therefore let $s = mi+k+j$. This equation then becomes

$$\bar{\nu}_n' = \frac{1}{m} \sum_{s=-\infty}^{+\infty} \sum_{j=0}^{m-1} \left(\frac{s}{m} + \frac{m-1-2j}{2m}\right)^n \frac{1}{m} f\left(\frac{s}{m}\right)$$

from which, by means of the binomial theorem, we obtain

$$\bar{\nu}_n' = \sum_{i=0}^{n} \binom{n}{i} \left\{ \sum_{j=0}^{m-1} \left(\frac{m-1-2j}{2m}\right)^i \frac{1}{m} \right\} \left\{ \sum_{s=-\infty}^{+\infty} \left(\frac{s}{m}\right)^{n-i} f\left(\frac{s}{m}\right) \frac{1}{m} \right\} .$$

But

$$\mu_{n-i}' = \sum_{s=-\infty}^{+\infty} \left(\frac{s}{m}\right)^{n-i} f\left(\frac{s}{m}\right) \frac{1}{m} .$$

We therefore have

$$(1) \qquad \bar{\nu}_n' = \sum_{i=0}^{n} \binom{n}{i} b_i(m) \mu_{n-i}' ,$$

where

$$(2) \qquad b_i(m) = \sum_{j=0}^{m-1} \left(\frac{m-1-2j}{2m}\right)^i \frac{1}{m} .$$

We shall sometimes write $b_i$ instead of $b_i(m)$ in order to simplify the expression of an equation The change of order of summation

is based on the assumption that the $m$ summations $V_n'(t)$ converge absolutely, an assumption equivalent to that of the existence of $\mu_n'$ since $f(x)$ has only positive or zero values.

We see immediately from (2) that $b_{2k+1}(m) = 0$, since the terms of the summation cancel each other in pairs, with the possible addition of a middle term equal to zero. The calculation of $b_{2k}(m)$ may readily be effected by means of the Euler-Maclaurin sum formula[6]

$$\sum_{j=0}^{m-1} g(j+\tfrac{1}{2}) = \int_0^m g(t)\,dt + \sum_{i=1}^{\infty} \left[\frac{D_{2i}}{4^i(2i)!}\, g^{(2i-1)}(t)\right]_0^m ,$$

where

$$D_0 = 1,$$
$$D_2 = -\frac{1}{3},$$
$$D_4 = \frac{7}{15},$$
$$D_6 = -\frac{31}{21},$$
$$D_8 = \frac{127}{15}.$$

We substitute

$$g(t) = \frac{1}{m^{2k+1}}\left(t - \tfrac{1}{2}m\right)^{2k}$$

---

[6] See for example Norlund's Differenzenrechnung, Berlin 1924, especially formulae (39), p 27; (42), p. 28, and (49), p. 30. Formula (39) is

$$\sum_{i=0}^{n}\binom{n}{i}D_{n-i} - \sum_{i=0}^{n}(-1)^i\binom{n}{i}D_{n-i} = 0 \quad \text{for } n>1,\ D_0=1.$$

From this we may obtain the values of $D_{2i}$ and show that $D_{2i+1}=0$; also we obtain $\sum_{i=0}^{n}\binom{2n+1}{2i}D_{2i} = 0$ for $n>0$ which we shall employ in the proof of our formula (7).

obtaining

$$(3) \qquad b_{2k}(m) = \frac{1}{4^k(2k+1)} \sum_{i=0}^{k} \binom{2k+1}{2i} \frac{D_{2i}}{m^{2i}}$$

The first few values are:

$$b_0(m) = 1,$$

$$b_2(m) = \frac{1}{12}\left(1 - \frac{1}{m^2}\right),$$

$$b_4(m) = \frac{1}{240}\left(3 - \frac{10}{m^2} + \frac{7}{m^4}\right),$$

$$b_6(m) = \frac{1}{1344}\left(3 - \frac{21}{m^2} + \frac{49}{m^4} - \frac{31}{m^6}\right),$$

$$b_8(m) = \frac{1}{11520}\left(5 - \frac{60}{m^2} + \frac{294}{m^4} - \frac{620}{m^6} + \frac{381}{m^8}\right).$$

A control on the values of $b_i(m)$ may be obtained by substituting $m = 1$ ; then all except $b_0$ vanish as $b_i(1) = 0$, for $i > 0$.

Having in (1) and (3) obtained the expression of the average grouped moment of a discrete distribution in terms of the true moments, we wish to solve for the true moments in terms of the average grouped moments. We shall obtain this solution by the method of undetermined coefficients. Let

$$(4) \qquad \mu_n' = \sum_{j=0}^{n} \binom{n}{j} A_{n-j} \bar{\nu}_j' .$$

Substituting this in (1) we shall have

$$\bar{\nu}_n' = \sum_{i=0}^{n} \binom{n}{i} b_i \sum_{j=0}^{n-i} \binom{n-i}{j} A_{n-i-j} \bar{\nu}_j' ,$$

from which we obtain

$$\bar{\nabla}_n' = \sum_{j=0}^{n} \binom{n}{j} \bar{\nabla}_j' \sum_{i=0}^{n-j} \binom{n-j}{i} b_i A_{n-j-i} \quad ,$$

by a change of order of summation effected by applying the Dirchlet sum·formula[7]

$$\sum_{i=0}^{n} \sum_{j=0}^{n-i} w(i,j) = \sum_{j=0}^{n} \sum_{i=0}^{n-j} w(i,j).$$

Equating coefficients of $\bar{\nabla}_j'$ gives us the recurring formula

$$\sum_{i=0}^{k} \binom{k}{i} b_i A_{k-i} = 0 \text{ for } k > 0,$$

together with the initial condition $A_0 = 1$. This may also be written

$$(5) \qquad A_k = -\sum_{i=0}^{k-1} \binom{k}{i} b_{k-i} A_i, \text{for } k \geq 1; \; A_0 = 1.$$

Ordinarily in an expression such as (4) we would have written $A_{n-j}(n)$ instead of $A_{n-j}$ ; had we done so in this case, we would now drop the functional expression as we have shown that the value of $A_{n-j}(n)$ depending only on $n-j$ is completely independent of $n$ . The coefficients $A_{n-j}$ are also independent of the position of the origin since if in

$$\mu'_{n:x+h} = \sum_{i=0}^{n} \binom{n}{i} h^i \mu'_{n-i.x} ,$$

[7] For the method of derivation of this formula see, for example, Steffensen's Interpolation (Baltimore, 1927), p. 91-92.

we substitute

$$\mu'_{n-i\cdot x} = \sum_{j=0}^{n-i} \binom{n-i}{j} A_j \bar{\nu}'_{n-i-j\cdot x},$$

we shall have

$$\mu'_{n.x+h} = \sum_{j=0}^{n} \binom{n}{j} A_j \sum_{i=0}^{n-j} \binom{n-j}{i} h^i \bar{\nu}'_{n-j-i.x},$$

and hence

$$\mu'_{n:x+h} = \sum_{j=0}^{n} \binom{n}{j} A_j \bar{\nu}'_{n-j\cdot x+h}.$$

If in (5) we substitute $k = 1$ we shall obtain $A_1 = -b_1 = 0$. Moreover in general $A_{2i+1} = 0$, since by induction if

$$A_1 = A_3 = \cdots = A_{2i-1} = 0,$$

the terms of the summation (5) will have respectively the zero factors

$$b_{2i+1},\ A_1,\ b_{2i-1},\ A_3, \cdots\ b_3,\ A_{2i-1},\ b_1.$$

Also from (5) we obtain:

$$\begin{aligned}
&A_0 = 1,\\
&A_2 = -b_2\\
&A_4 = -b_4 + 6(b_2)^2,\\
&A_6 = -b_6 + 30 b_2 b_4 - 90 (b_2)^3,\\
&A_8 = -b_8 + 56 b_2 b_6 + 70 (b_4)^2 - 1260 (b_2)^2 b_4 + 2520 (b_2)^4.
\end{aligned}$$

An observation of these expressions of the A's suggests the formula:

$$A_{2i} = \sum \frac{(-1)^k (2i)!\, k!\, (b_2)^{a_1} (b_4)^{a_4} \cdots (b_{2j})^{a_j}}{(2!)^{a_1} (4!)^{a_2} \cdots ((2j)!)^{a_j} (a_1!)(a_2!) \cdots (a_j!)}$$

where $k = a_1 + a_2 \cdots a_j$, the summation extending over all positive integral or zero values of $a_1, a_2, \cdots\cdots a_j$ satisfying $a_1 + 2a_2 + \cdots + j a_j = i$. That this formula holds in general may be proved by induction: assume it true for $i = 0, 1, \cdots j-1$ and substitute in (5) for $k = 2j$. Upon collecting terms according to products of the b's we shall have established this formula also for $i = j$, and hence for every positive integral value.

If in the expressions of the A's in terms of the b's we substitute the values of the b's in terms of m, we shall obtain the expression of the A's in terms of m. Thus we have

$$\begin{aligned} &A_0 = 1, \\ &A_2 = -\frac{1}{12}\left(1 - \frac{1}{m^2}\right), \\ (6)\quad &A_4 = \frac{1}{240}\left(7 - \frac{10}{m^2} + \frac{3}{m^4}\right), \\ &A_6 = -\frac{1}{1344}\left(31 - \frac{49}{m^2} + \frac{21}{m^4} - \frac{3}{m^6}\right), \\ &A_8 = \frac{1}{11520}\left(381 - \frac{620}{m^2} + \frac{294}{m^4} - \frac{60}{m^6} + \frac{5}{m^8}\right). \end{aligned}$$

A comparison of the values of $A_0$ with $b_0$, of $A_2$ with $b_2$, of $A_4$ with $b_4$, of $A_6$ with $b_6$, and of $A_8$ with $b_8$ shows a remarkable similarity between the coefficients in $A_{2i}$ and those in $b_{2i}$; in fact we observe that $A_{2i} = \frac{1}{m^{2i}} b_{2i}\left(\frac{1}{m}\right)$. Substituting $\frac{1}{m}$ for $m$ in (3) and dividing by $m^{2i}$, we obtain

$$(7) \qquad A_{2k}(m) = \frac{1}{4^k(2k+1)} \sum_{i=0}^{k} \binom{2k+1}{2i+1} \frac{D_{2k-2i}}{m^{2i}}$$

In order to prove that (7) is true in general, we assume it true up to a certain point and prove it true for the next highest value of $k$. That is we assume

$$A_{2i} = \frac{1}{m^{2i}} b_{2i}\left(\frac{1}{m}\right) \text{ for } i = 0, 1, \quad k-1$$

and substitute in

$$A_{2k} = -\sum_{i=0}^{k-1} \binom{2k}{2i} b_{2k-2i} A_{2i},$$

another form of (5) since $b_{2k-2i-1} = 0$. From (3) we have

$$b_{2k-2i} = \frac{1}{4^{k-i}(2k-2i+1)} \sum_{j=0}^{k-i} \binom{2k-2i+1}{2j} \frac{D_{2j}}{m^{2j}},$$

and

$$A_{2i} = \frac{1}{4^i(2i+1)} \sum_{r=0}^{i} \binom{2i+1}{2r} \frac{D_{2r}}{m^{2i-2r}}$$

After this substitution we arrange the terms of $A_{2k}$ according to powers of $\frac{1}{m}$ obtaining

$$A_{2k} = \frac{1}{4^k(2k+1)} \sum_{s=0}^{k} \binom{2k+1}{2s+1} \frac{1}{m^{2s}} \left\{ \sum_j \binom{2s+1}{2j} D_{2j} \right\}$$

$$\cdot \left\{ -\frac{1}{(2k-2s+1)} \sum_r \binom{2k-2s+1}{2r} D_{2r} \right\},$$

where $s = i - r + j$. When $s = 0, 1, \quad k-1$ the summation extends from $r = 0$ to $r = k - s$ for $j \neq 0$, but from $r = 0$ to

$r = k-s-1$ for $j=0$. Since

$$\sum_{r=0}^{k-s}\binom{2k-2s+1}{2r}D_{2r}=0,$$

for $s<k$, the summation as to $r$ gives zero for $j \neq 0$ but

$$-(2k-2s+1)D_{2k-2s}, \quad \text{for } j=0.$$

At the same time for $j=0$, the factor $\binom{2s+1}{2j}D_{2j}$ equals unity and we have the desired terms for $s=0, 1, \cdot \cdots \cdot \cdot k-1$. For $s=k$ we have constantly $r=0$, the summation as to $j$ being from $j=1$ to $j=k$; we therefore have

$$\sum_{j=1}^{s}\binom{2s+1}{2j}D_{2j}=-D_0,$$

at the same time that

$$-\frac{1}{(2k-2s+1)}\sum_{r=0}^{0}\binom{2k-2s+1}{2r}D_{2r}=-1.$$

We therefore come again to formula (7) with $i$ replaced by $s$. Hence formula (7) is true for every positive integral value of $k$. The first few values of the A's have been calculated in (6), others may be easily obtained by substituting the value of the Eulerian numbers from some table of $D_{2i}$.[8]

Formulae (4) and (7) give us the expression of the true moment

$$\mu'_n=\sum_{j=-\infty}^{+\infty}\left(\frac{j}{m}\right)^n f\left(\frac{j}{m}\right),$$

of a discreet distribution in terms of the set of average grouped moments

[8] Nörlund, loc. cit., Tafel 4, p 458, gives the value up to $D_{20}$

$$\bar{\nu}_i' = \sum_{j=-\infty}^{+\infty} \frac{1}{m}\left(\frac{j}{m}+\frac{m-1}{2m}\right)^i \phi\left(\frac{j}{m}\right)+\frac{m-1}{2m}\bigg).$$

Employing the particular values given in (6), we have the formula

$$\begin{aligned}\mu_n' = \bar{\nu}_n' &- \frac{1}{12}\binom{n}{2}\left(1-\frac{1}{m^2}\right)\bar{\nu}_{n-2}' + \frac{1}{240}\binom{n}{4}\left(7-\frac{10}{m^2}+\frac{3}{m^4}\right)\bar{\nu}_{n-4}' \\ &- \frac{1}{1344}\binom{n}{6}\left(31-\frac{49}{m^2}+\frac{21}{m^4}-\frac{3}{m^6}\right)\bar{\nu}_{n-6}' \\ &+ \frac{1}{11520}\binom{n}{8}\left(381-\frac{620}{m^2}+\frac{294}{m^4}-\frac{60}{m^6}+\frac{5}{m^8}\right)\bar{\nu}_{n-8}' \end{aligned} \tag{8}$$

For any particular value of $n$, this series terminates and we may therefore apply the ordinary theory of limits to (8). Thus we obtain

$$\begin{aligned}\mu_n' = \bar{\nu}_n' &- \frac{1}{12}\binom{n}{2}\bar{\nu}_{n-2}' + \frac{7}{240}\binom{n}{4}\bar{\nu}_{n-4}' \\ &- \frac{31}{1344}\binom{n}{6}\bar{\nu}_{n-6}' + \frac{127}{3840}\binom{n}{8}\bar{\nu}_{n-8}' - \cdots,\end{aligned} \tag{9}$$

the expression of the true moment $\mu_n' = \int_{-\infty}^{+\infty} x^n f(x)\,dx$ of a continuous distribution in terms of the set of average grouped moments

$$\bar{\nu}_i' = \int_{-\infty}^{+\infty} \left(x+\tfrac{1}{2}\right)^i \phi\left(x+\tfrac{1}{2}\right)dx = \int_{-\infty}^{+\infty} x^i \phi(x)\,dx.$$

We have thus completely solved the first of our two problems; we have obtained the expression of the true moments in terms of the average grouped moments without any assumption other than the existence of $\mu_n'$. The existence of $\mu_n'$ requires the convergence

of the summation or integration as the lower limit approaches $-\infty$ and as the upper limit approaches $+\infty$ independently.

If in (8) we replace

$$\bar{\nu}_i' = \sum_{j=-\infty}^{+\infty} \frac{1}{m}\left(\frac{j}{m}+\frac{m-1}{2m}\right)^i \phi\left(\frac{j}{m}+\frac{m-1}{2m}\right)$$

by

$$\nu_i' = \sum_{j=-\infty}^{+\infty} (x_j)^i \phi(x_j),$$

we will have the general Sheppard-Carver formula. Since there is no approximation involved in (8), any error in the Sheppard-Carver formulae must be a result of the error involved in replacing the average grouped moments $\bar{\nu}_i'$ by the raw moments $\nu_i'$. By definition $\bar{\nu}_i'$ is the average of the

$$\bar{\nu}_i'(t) = \sum_{j=-\infty}^{+\infty} (j+t)^i \phi(j+t),$$

and, therefore, if we take any particular grouping at random,

$$\nu_i' = \sum_{j=-\infty}^{+\infty} (x_j)^i \phi(x_j),$$

is the mean of a random sample of one from the parent distribution $\nu_i'(t)$ and hence the most probable value of $\bar{\nu}_i'$. The Sheppard-Carver formula, therefore, gives the most probable value of the true moment $\mu_n'$ of a discrete distribution in the sense that these formulae eliminate the systematic errors due to grouping.

Similarly, we shall obtain Sheppard's corrections if in (9) we

replace

$$\bar{\nu}_i' = \int_{-\infty}^{+\infty} x^i \phi(x)\,dx,$$

by

$$\nu_i' = \sum_{j=-\infty}^{+\infty} (x_j)^i \phi(x_j).$$

These formulae give the most probable value of the true moments $\mu_n'$ for a continuous distribution in the same sense as do the Sheppard-Carver formulae for a discrete distribution.

The Sheppard corrections for continuous distributions and the Sheppard-Carver corrections for discrete distributions give the most probable value of the true moments $\{\mu_n'\}$ of a distribution $f(x)$ in the sense that they give an approximate value for $\mu_n'$ which is correct *on the average*. That is these formulae eliminate the systematic errors due to grouping whatever the distributional function $f(x)$ so long as the moments under consideration exist. While it is true that the accidental errors not accounted for in these corrections may not be negligible, these formulae do give the most probable value of $\mu_n'$ for a particular grouping and hence have a basis for universal application.

John R. Abernethy

# ON MULTIPLE AND PARTIAL CORRELATION COEFFICIENTS OF A CERTAIN SEQUENCE OF SUMS

*By*
CARL H. FISCHER

In a recent paper* the writer considered a sequence of $q$ variables defined as follows: The first variable, $x_1$, is defined as the sum of $n_1$ values of a variable, $t$, drawn at random from a population characterized by a rather arbitrary continuous probability function, $f(t)$. Each succeeding variable, $x_i$, $(i > 1)$, is defined as the sum of $k_{i-1,i}$ values of $t$ drawn at random from the $n_{i-1}$ values composing $x_{i-1}$, plus the sum of $n_i - k_{i-1,i}$ values of $t$ drawn at random from the parent population.

For variables thus defined, it was proved that the correlation coefficient between any two consecutive sums, $x_i$ and $x_{i+1}$, is independent of the probability function, $f(t)$, and is given by

$$r_{x_i x_{i+1}} = \frac{k_{i,i+1}}{(n_i n_{i+1})^{1/2}} \quad . \tag{1}$$

It was further shown that the correlation coefficient between two variables not consecutive in the sequence is equal to the product of the respective coefficients of correlation between all intermediate pairs of consecutive variables. Thus, the coefficient of correlation between $x_j$ and $x_p$, $(j < p)$, is

$$r_{x_j x_p} = r_{x_j x_{j+1}} \cdot r_{x_{j+1}, {j+2}} \cdots r_{x_{p-2}, p-1} \cdot r_{x_{p-1} x_p} \ ;$$

or, in a simpler notation,

$$r_{jp} = r_{j,j+1} \cdot r_{j+1,j+2} \cdots r_{p-2,p-1} \cdot r_{p-1,p} \ . \tag{2}$$

---

* On Correlation Surfaces of Sums with a Certain Number of Random Elements in Common. Annals of Mathematical Statistics, Vol. IV, pp. 103-126. May, 1933

Let us now determine the multiple and partial correlation co-efficients existing among a sequence of variables thus defined.

Consider the fundamental symmetric determinant, $R$ , which, with its various co-factors, appears in the standard formulas for multiple and partial correlation coefficients.[1] If we substitute for each $r_{ij}$ , $(j \neq i-1, i, i+1)$, in $R$ , its equivalent from equation (2), we have

$$(3)\ R = \begin{vmatrix} 1 & r_{12} & r_{12}r_{23} & \cdots & r_{12}r_{23}\cdots r_{q-1,q} \\ r_{12} & 1 & r_{23} & \cdots & r_{23}r_{34}\cdots r_{q-1,q} \\ r_{12}r_{23} & r_{23} & 1 & \cdots & r_{34}\cdots r_{q-1,q} \\ r_{12}r_{23}r_{34} & r_{23}r_{34} & r_{34} & \cdots & r_{45}\cdots r_{q-1,q} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{12}r_{23}\cdots r_{q-2,q-1} & r_{23}\cdots r_{q-2,q-1} & r_{34}\cdots r_{q-2,q-1} & \cdots & r_{q-1,q} \\ r_{12}r_{23}\cdots r_{q-1,q} & r_{23}\cdots r_{q-1,q} & r_{34}\cdots r_{q-1,q} & \cdots & 1 \end{vmatrix}$$

Multiply the second row of $R$ by $r_{12}$ and subtract it from the first row. Now multiply the third row by $r_{23}$ and subtract it from the second row Continue this process, multiplying the $j$-th row by $r_{j-1,j}$ and subtracting this row from the $(j-1)$st row, until all possible rows have been so treated. Equation (3) may now be written

---

[1] H. L. Rietz, "Mathematical Statistics", Carus Monograph No. 4, pp. 94-100.

$$
(4)\ R = \begin{vmatrix}
(1-r_{12}^2) & 0 & 0 & 0 & 0 \\
r_{12}(1-r_{23}^2) & (1-r_{23}^2) & 0 & 0 & 0 \\
r_{12}r_{23}(1-r_{34}^2) & r_{23}(1-r_{34}^2) & (1-r_{34}^2)\cdot\ \cdot & 0 & 0 \\
r_{12}r_{23}r_{34}(1-r_{45}^2) & r_{23}r_{34}(1-r_{45}^2) & r_{34}(1-r_{45}^2) & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
r_{12}r_{23}\cdot r_{q-2,q-1}(1-r_{q-1,q}^2) & r_{23}\ r_{q-2,q-1}(1-r_{q-1,q}^2) & r_{34}\ r_{q-2,q-1}(1-r_{q-1,q}^2)\cdot & (1-r_{q-1,q}^2) & 0 \\
r_{12}r_{23}\ r_{q-1,q} & r_{23}\ r_{q-1,q} & r_{34}\ r_{q-1,q} & r_{q-1,q} & 1
\end{vmatrix}
$$

The expansion of $R$ may now be readily accomplished. The application of this same method of procedure to each of the various $R_{ij}$, (where $R_{ij}$ is the co-factor of the element $r_{ij}$), yields without difficulty the results made use of in the remainder of this paper.

*A. Multiple Correlation Coefficients.*

The formula for the multiple correlation coefficient of one variable on the remaining $q-1$ variables is

$$
(5) \qquad r_{j\cdot 123\cdots j-1,\ j+1,\ \cdot\ \ q} = \left(\frac{1-R}{R_{jj}}\right)^{\frac{1}{2}} .
$$

From equations (3) and (4) we derive the following expressions for the necessary $R$ and $R_{jj}$, $(j = 1, 2, 3, \cdots q)$.

$$R=(1-r_{12}^2)(1-r_{23}^2)(1-r_{34}^2)(1-r_{45}^2)\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$R_{11}=\qquad(1-r_{23}^2)(1-r_{34}^2)(1-r_{45}^2)\quad\cdot(1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$R_{22}=\qquad(1-r_{12}^2r_{23}^2)(1-r_{34}^2)(1-r_{45}^2)\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$R_{33}=(1-r_{12}^2)(1-r_{23}^2r_{34}^2)(1-r_{45}^2)\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q});\tag{6}$$

$$\vdots$$

$$R_{jj}=(1-r_{12}^2)(1-r_{23}^2)\cdot\ \cdot(1-r^2_{j-2,j-1})(1-r^2_{j-1,j}\,r^2_{j,j+1})(1-r^2_{j+1,j+2})\cdots$$

$$\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q});$$

$$\vdots$$

$$R_{q-1,q-1}=(1-r_{12}^2)(1-r_{23}^2)\cdots\cdot(1-r^2_{q-3,q-2})(1-r^2_{q-2,q-1}r^2_{q-1,q});$$

$$R_{qq}=(1-r_{12}^2)(1-r_{23}^2)\quad\cdot\quad\cdots(1-r^2_{q-3,q-2})(1-r^2_{q-2,q-1}).$$

Upon substituting the proper values from (6) in formula (5) for the multiple correlation coefficients of the first and last variables, respectively, on the others in the sequence, we find

$$r_{1\cdot 234\ \ q}=r_{12};\tag{7}$$

$$r_{q\cdot 123\cdot\ \cdot q-1}=r_{q-1,q}\tag{8}$$

The multiple correlation coefficient of any other variable, $x_j$ , on the remaining $q-1$ variables is given by

$$(9)\qquad \begin{aligned} r_{j\cdot 123\cdots q} &= \left[\frac{1-(1-r^2_{j-1,j})(1-r^2_{j,j+1})}{(1-r^2_{j-1,j}r^2_{j,j+1})}\right]^{\frac{1}{2}} \\ &= \left[\frac{(r^2_{j-1,j}+r^2_{j,j+1}-2r^2_{j-1,j}\,r^2_{j,j+1})}{(1-r^2_{j-1,j}\,r^2_{j,j+1})}\right]^{\frac{1}{2}}. \end{aligned}$$

It is to be noted that the right member of (9) is independent of all of the simple correlation coefficients except $r_{j-1,j}$ and $r_{j,j+1}$ .

*B. Partial Correlation Coefficients.*

The formula for the partial correlation coefficient between any two variables is

$$(10)\qquad r_{ij\cdot 1234\cdots q} = \frac{-R_{ij}}{(R_{ii}\,R_{jj})^{1/2}} .$$

From equation (3) we derive the following expressions for the co-factors of elements other than those of the principal diagonal of the determinant. Because of the symmetry of the fundamental determinant, we know that $R_{ij} = R_{ji}$ ; hence in expanding the co-factors of the elements of each row, we shall consider only the $R_{ij}$ where $i \leqq j$ .

1. The co-factors of the elements of the first row are

$$(11)\qquad \begin{aligned} -R_{12} &= r_{12}\,(1-r^2_{23})(1-r^2_{34})\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q}); \\ R_{1i} &= 0,\ (i = 3,4,5,\cdots q). \end{aligned}$$

2. The co-factors of the elements of the second row are

$$(12)\qquad \begin{aligned} -R_{23} &= (1-r^2_{12})\,r_{23}\,(1-r^2_{34})\cdots(1-r^2_{q-2,q-1})(1-r^2_{q-1,q}), \\ R_{2i} &= 0,\ \ (i = 4,5,6,\cdots q). \end{aligned}$$

3. The co-factors of the elements of the j-th row are

$$(13)\quad -R_{j,j+1} = (1-r_{12}^2)(1-r_{23}^2)\cdots(1-r_{j-1,j}^2)r_{j,j+1}(1-r_{j+1,j+2}^2)\cdot(1-r_{q-1,q}^2);$$
$$R_{ji} = 0,\ (i = j+2,\ j+3,\ \cdots\cdots q)$$

4. The co-factor of the last element of the $q$-th row is

$$(14)\qquad -R_{q-1,q} = (1-r_{12}^2)(1-r_{23}^2)\cdots(1-r_{q-2,q-1}^2)r_{q-1,q}\,.$$

We see at once that all partial correlation coefficients between non-consecutive variables vanish, as each co-factor $R_{ij} = 0$ if $i \neq j-1\ ,\ j\ ,\ j+1$ . The non-vanishing coefficients, those between consecutive variables, are given below.

$$(15)\qquad r_{12\cdot 345\cdot\cdot q} = r_{12}\left[\frac{(1-r_{23}^2)}{(1-r_{12}^2\,r_{23}^2)}\right]^{\frac{1}{2}}\ ;$$

$$(16)\qquad r_{j,j+1\cdot 1234\quad q} = r_{j,j+1}\left[\frac{(1-r_{j-1,j}^2)(1-r_{j+1,j+2}^2)}{(1-r_{j-1,j}^2\,r_{j,j+1}^2)(1-r_{j,j+1}^2 r_{j+1,j+2}^2)}\right]^{\frac{1}{2}};$$

$$(17)\qquad r_{q-1,q\cdot 1234\cdots q-2} = r_{q-1,q}\left[\frac{(1-r_{q-2,q-1}^2)}{(1-r_{q-2,q-1}^2\,r_{q-1,q}^2)}\right]^{\frac{1}{2}}\ .$$

From (16) we can state that in general the partial correlation coefficient of consecutive sums $x_j$ and $x_{j+1}$ is independent of all simple correlation coefficients except $r_{j-1,j}$, $r_{j,j+1}$, and $r_{j+1,j+2}$.

## C. *Summary*

To summarize, we have shown that

1. The multiple correlation coefficient of a variable $x_j$ on the remaining variables of our sequence is independent of all simple correlation coefficients except those between $x_j$ and the immediately preceding and the immediately following variables, respectively
2. The partial correlation coefficients between all pairs of non-consecutive variables in our sequence are zero; a result that appeals to the intuition when it is recalled that we are eliminating the effect of the variables that form the connecting links between the two under consideration.
3. The partial correlation coefficient of any pair of consecutive variables, $x_j$ and $x_{j+1}$, is independent of all simple correlation coefficients except those between the two consecutive variables in question, between the first of these and the variable immediately preceding it, and between the second of the pair and the variable immediately following it.

Carl H. Fischer

# AN EXPERIMENT REGARDING THE $\chi^2$ TEST

By
SELBY ROBINSON*

## 1. *Introduction*

R. A. Fisher has proposed that in case the hypothesis being tested has been partially obtained from the data, the Elderton table[1] for $\chi^2$ should be entered with $n'$ equal to, not the number of frequency classes, but the number of frequency classes minus the number of statistics computed from the data. It has been proved under certain restrictions that this theory holds in the limit as the size of the sample approaches infinity.[2]

For samples of moderate size our only guide is experimental evidence, which indicates that Fisher's method is satisfactory in practice.[3] This is true in particular of the evidence presented in the present paper which describes a coin tossing experiment suggested by Professor H. L Rietz.

## 2. *The Experiment*

The experimental work here considered was done by seventy students each of whom tossed seven coins 128 times. In any one of the seventy experiments, this results in a frequency distribution of 128 items divided into eight frequency classes But we lumped together the classes of zero heads and one head and likewise for six heads and seven heads, so that we had six frequency classes. If for every coin on every throw the probability of heads

---

* National Research Fellow.

1 Karl Pearson, Tables for statisticians and biometricians, (1914), Table XII.

2 J. Neyman and E. S. Pearson, Biometrika, V. 20A (1928), pp. 263-294

3 Yule, Journal of the Royal Statistical Society, V. 85, pp. 95-104; Brownlee, *ibid.*, V 87, p 76; Neynam and Pearson, Loc. Cit.; Sheppard, Phil. Trans., A. V., 228, p. 115.

is one-half, the expected numbers $\tilde{m}_i$ in the six frequency classes are 8, 21, 35, 35, 21, 8. The divergence of the actual distribution $n_1$ , $n_2$ , . . ., $n_6$ , from the theoretical one is measured by

$$\chi^2 = \sum_{i=1}^{i=6} \frac{(n_i - \tilde{m}_i)}{\tilde{m}_i} .$$

If we had possessed only one sample of 128 throws, we would have looked in Elderton's table with the value of $\chi^2$ and with $n' = 6$ (the number of frequency classes) to find the probability $P_\chi$ that a sample would by chance deviate from the expected distribution so much as the actual sample had deviated. But having seventy samples, we compared the distribution of our seventy values of $\chi^2$ with that expected from $n = 6$ . The arithmetic mean of our seventy values of $\chi^2$ was 4.62 whereas the expected value was five; a deviation which could very well occur by chance. So our results are consistent with the hypothesis that the probability of a head is always one-half.

We considered next the following composite hypothesis: the probability $p$ of heads is the same for all coins on all throws. For any sample of 128 throws, we took as the estimate of $p$ the actual proportion of heads among the 7 x 128 possibilities. From this value we calculated the expected frequencies, $m_1$, . . ., $m_6$. For each of the seventy samples, we computed

$$\chi_1^2 = \sum_{i=1}^{i=6} \frac{(n_i - m_i)^2}{m_i} .$$

When we used the $\chi^2$ test to compare this distribution of $\chi_1^2$s with that expected for $n' = 6$ , we found that $P_\chi = .01$. But when we compared our distribution with that expected for $n' = 6 - 1 = 5$ we found that $P_\chi = .82$. The mean value of our values of $\chi_1^2$ was 3.97 compared with $6 - 1 - 1 = 4$ demanded by Fisher's theory, and with five by the theory that $\chi_1^2$ is distributed as

$\chi^2$. If the latter theory were correct the probability of the mean of seventy values of $\chi_1^2$ being so far away from five, is 007. That our distribution of $\chi_1^2$ corresponds to that expected for $n' = 5$ whereas our distribution of the values of $\chi^2$ corresponds to $n' = 6$, can be seen from the following table.[4]

| Values of $\chi^2$ (or $\chi_1^2$). | Observed frequency of $\chi^2$ | Expected frequency $n'=6$ | Observed frequency of $\chi_1^2$ | Expected frequency $n'=5$ |
|---|---|---|---|---|
| 0—1 | 3 | 2 6 | 7 | 6 3 |
| 1—2 | 8 | 8 0 | 12 | 12.2 |
| 2—3 | 11 | 10 4 | 14 | 12.5 |
| 3—4 | 14 | 10 5 | 12 | 10.6 |
| 4—5 | 12 | 9.3 | 10 | 8.3 |
| 5—7 | 11 | 13 7 | 7 | 10.6 |
| 7—9 | 5 | 7.8 | 3 | 5.2 |
| greater than 9 | 6 | 7.6 | 5 | 4.3 |

[4] In computing $P_\chi$ we combined the first two classes of this table and also the last two, thus making $n' = 6$.

Selby Robinson

# ON SAMPLING FROM COMPOUND POPULATIONS*

*By*
GEORGE MIDDLETON BROWN

## *Introduction.*

The decided asymmetry or the multimodality of certain frequency distributions may have prompted the idea of the possibility of the existence of frequency curves, apparently single in character, but which, on further investigation, might be shown to be actually composite. In other words, apparently homogeneous material may prove to be heterogeneous, or divisible into two or more distinct homogeneous groups

The above ideas lead naturally to the problem of dissecting a compound frequency function into its various components. Karl Pearson[1] successfully solved such a problem, using the method of moments, on the assumption that the compound parent population was composed of two normal components Each component curve has three parameters, the mean (or position of axis), the standard deviation, and the area (or total frequency). One requires therefore, six relations between the parameters of the given compound frequency curve, and those of its two components, in order to determine six unknowns. The ultimate solution of the problem turns on the determination of the zeros of a nonic equation, the location of whose real roots is obtained, to successive approximations, by means of the so-called Sturm's functions.

The dissection problem was taken up later, first in a paper by Charlier,[2] then in a joint paper by Charlier and Wicksell[3] who

---

* A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Science in the University of Michigan. June, 1933

[1] On the dissection of frequency curves into normal curves Karl Pearson. Phil. Trans. Roy. Soc. Lond Vol 185, Pt. 1, pp 71-110. 1894A.

[2] Researches into the theory of probability. C. V. L. Charlier. Meddelanden frau Lunds Astron Observ. Sec 2. Bd 1 1906

[3] On the dissection of frequency functions. C. V. L. Charlier, and S D Wicksell. Arkiv. fur Matematik. Astron. och Fysik. (Meddelande) Band 18. No 6 1923.

considerably simplified the theory, finally arriving, however, at the fundamental nonic due to Pearson, for the solution of which, they suggested the use of a graphical method. They also studied special cases of the more general problem, e g the means of the two components assumed known, the compound curve assumed symmetrical, or the standard deviations of the two components supposed equal. In addition, they extended the problem to the case of frequency functions of two variates.

In the present paper, I propose to investigate the sampling problem in the case of compound distribution functions, and from a consideration of the dissection problem, one is led to a division of the present investigation into two main parts, for the following reasons.

On the one hand, in sampling from a compound population, if we do not know the proportion contributed to the total frequency of the sample by each of the two components of the parent population, we are essentially sampling from a single population. That is, random samples of $N$ are drawn from a single composite parent population made up of two components Hence, the previously obtained results for sampling from a single parent population will be available if we derive expressions for the parameters of the compound parent in terms of the parameters of its components. This is done in Part 1.

On the other hand, however, if we know the proportion contributed to the total frequency of a sample by each of the two components, the situation differs entirely from that studied in Part 1. Here we are concerned with sampling from two distinct parent populations, and in Part 2, I develop a method for dealing with this problem. Thus, in Part 2, it is assumed that samples of $r$ and $s$ respectively are drawn from two distinct parent populations, and these two samples are then combined to yield a sample of $r+s=N$ from the combined populations

Therefore in Part 1, we are essentially sampling from a single parent population, whereas in Part 2, we are sampling from multiple populations.

The developments of Part 2 yield some new sampling results for sampling from two parent populations. In Section 6, I derive expressions for the semi-invariants "of moments about a fixed point" in samples from the compound frequency function, in terms of the corresponding semi-invariants of the moments of its components. In Section 7, expressions are derived for the semi-invariants of "moments in samples from the compound population about the mean of the combined sample," in terms of $r$ and $s$, and the semi-invariants of the two components themselves.

The occurrence of a certain class of well-known polynomials in the development of Section 1, is of especial interest, since these are, except perhaps for sign, the semi-invariants of the binomial distribution, and have some rather important properties, and their further study, although not pertinent to the problem in hand, should yield some very interesting results.

Section 5 is devoted to the discussion of the case in which a limiting compound frequency function exists, under certain assumptions regarding the nature of its components, where the number of the latter is allowed to increase indefinitely. This idea of a limit frequency function would appear to indicate the possibility of a new approach to the theory of frequency curves, in which the variable may now be a complete frequency distribution in itself.

This investigation was begun on the suggestion of Professor C. C. Craig, of the University of Michigan, U. S. A. to whom I am indebted for constant inspiration and guidance during its pursuit.

PART 1

*Section 1. The semi-invariants of the compound frequency function, in terms of the semi-invariants of its two normal components.*

The main object of this first section, is to obtain expressions for the parameters of a compound population in terms of the parameters of its two normal components, and to this end, I shall use the following definition of the semi-invariants of Thiele.[1]

$$e^{\lambda_1 t+\lambda_2 \frac{t^2}{2!}+\lambda_3\frac{t^3}{3!}+\cdots} = \int_{-\infty}^{\infty} f(x)\cdot e^{xt}\,dx.$$

I write therefore

$$f(x) = p\ \varphi_1(x) + q\ \varphi_2(x).$$

in which $f(x)$ is the compound frequency function, $\varphi_1(x), \varphi_2(x)$ are its two normal components, and $p+q=1$

If $L_1$, $L_2$, etc. are the semi-invariants of $f(x)$, then

$$(1)\ e^{L_1 t+L_2\frac{t^2}{2!}+L_3\frac{t^3}{3!}+\cdots} = p e^{m_1 t+\sigma_1^2\frac{t^2}{2!}}\left\{1+\frac{q}{p}\cdot e^{(m_2-m_1)t+(\sigma_2^2-\sigma_1^2)\frac{t^2}{2!}}\right\}$$

where $m_1, m_2, \sigma_1, \sigma_2$, are the means and standard deviations of $\varphi_1(x), \varphi_2(x)$ respectively. For convenience, I write

$$m_2 - m_1 = a. \qquad \sigma_2^2 - \sigma_1^2 = b. \qquad \frac{q}{p} = r$$

We wish to express the $L_n$ in (1) in terms of the quantities $m_1$, $m_2, \sigma_1, \sigma_2$, $p$, or $q$.

Taking logarithms in (1),

---

[1] Numerous references relating to the theory of semi-invariants may be found at the end of "An application of Thiele's semi-invariants to the sampling problem", C. C. Craig. Metron, Vol. 7, No 4, 1928, p. 73

$$(2)\quad \begin{aligned} L_1 t + L_2 \frac{t^2}{2!} + L_3 \frac{t^3}{3!} + \cdot \\ = \log p + \left(m_1 t + \sigma_1^2 \frac{t^2}{2!}\right) + \log\left(1 + re^{at + \frac{bt^2}{2!}}\right). \end{aligned}$$

We require now a suitable form for the expansion of the third term of the right member of (2) in successive powers of $t$ . We have

$$\log\left(1+re^{at+\frac{bt^2}{2}}\right) = \log(1+r) + \log\left\{1+q\left(e^{at+\frac{bt^2}{2}}-1\right)\right\}.$$

Further

$$(3)\quad \log\left(1+re^{at+\frac{bt^2}{2}}\right) = \sum_{j=1}^{\infty} q^j\left(e^{at+\frac{bt^2}{2}}-1\right)^j(-1)^{j+1}.$$

The complete representation of terms of the type $e^{i\left(at+\frac{bt^2}{2}\right)}$ in the right member of (3) will be

$$\sum_{k=1}^{\infty}\sum_{i=0}^{k} \frac{(-1)^{1+i}\cdot q^k}{k}\binom{k}{i} e^{i\left(at+\frac{bt^2}{2}\right)}.$$

But

$$(4)\quad e^{i\left(at+\frac{bt^2}{2}\right)} = \sum_{j=0}^{\infty} \frac{i^j}{j!}\left(at+\frac{bt^2}{2}\right)^j.$$

Therefore, the coefficient of $\frac{t^n}{n!}$ in the right member of (4) is

$$\sum_{j=\left[\frac{n+1}{2}\right]}^{n} \frac{n!\, a^{2j-n} \cdot b^{n-j} \cdot i^j}{2^{n-j} \cdot (2j-n)!\,(n-j)!}$$

where $\left[\frac{n+1}{2}\right]$ means the largest integer in $\frac{n+1}{2}$.

Then, the coefficient of $\frac{t^n}{n!}$ in the right member of (3) is

$$(5)\quad L_n = \sum_{k=1}^{n} \sum_{i=1}^{k} \frac{(-1)^{1+i} \cdot q^k}{k} \cdot \binom{k}{i} \sum_{j=\left[\frac{n+1}{2}\right]}^{n} \frac{n!\, a^{2j-n} \cdot b^{n-j} i^j}{2^{n-j}\,(2j-n)!\,(n-j)!}\,;\ n>2$$

and this is the relation sought, in which the semi-invariants of the compound frequency function are expressed in terms of the semi-invariants of its two normal components. Below, I have written out in detail the expressions for $L_1$ to $L_9$ inclusive.

$$L_1 = m_1 + aq.$$

$$L_2 = \sigma_1^2 + q\left[a^2(1-q) + b\right]$$

$$L_3 = a^3 pqq_3 + 3abpq.$$

$$(6)\qquad L_4 = a^4 pqq_4 + 6a^2 bpqq_3 + 3b^2 pq.$$

$$L_5 = a^5 pqq_5 + 10a^3 bpqq_4 + 15ab^2 pqq_3.$$

$$L_6 = a^6 pqq_6 + 15a^4 bpqq_5 + 45a^2b^2 pqq_4 + 15b^3 pqq_3$$

$$L_7 = a^7 pqq_7 + 21a^5 bpqq_6 + 105a^3b^2 pqq_5 + 105ab^3 pqq_4.$$

$$L_8 = a^8pqq_8 + 28a^6bpqq_7 + 210a^4b^2pqq_6 + 420a^2b^3pqq_5 + 105b^4pqq_4.$$

$$L_9 = a^9pqq_9 + 36a^7bpqq_8 + 378a^5b^2pqq_7 + 1260a^3b^3pqq_6 + 945ab^4pqq_5.$$

in which

$$q_3 = 1 - 2q.$$

$$q_4 = 1 - 6q + 6q^2.$$

$$q_5 = 1 - 14q + 36q^2 - 24q^3.$$

(7)

$$q_6 = 1 - 30q + 150q^2 - 240q^3 + 120q^4.$$

$$q_7 = 1 - 62q + 540q^2 - 1560q^3 + 1800q^4 - 720q^5.$$

$$q_8 = 1 - 126q + 1806q^2 - 8400q^3 + 16800q^4 - 15120q^5 + 5040q^6.$$

$$q_9 = 1 - 254q + 5796q^2 - 40824q^3 + 126000q^4 - 191520q^5 + 141120q^6 - 40320q^7.$$

The expressions for the $L_n$ in (6) have two properties, which enable one to write them down readily. In the first place, assuming that the polynomials in $q$ (or $p = 1 - q$) are suppressed, i.e. $q_2' = q_2 = pq$, $q_3' = pqq_3$, $q_4' = pqq_4$, etc., are all set equal to unity, then the resulting functions in "a" and "b" are readily obtained by means of

a well-known recursion formula. Secondly, considering the polynomials $q'_k$ as coefficients in the several terms of the original complete expressions for the $L_n$ in (6), for $n>2$, and arranging these expressions so that their corresponding terms appear in columns, the first terms in the first column, the second terms in the second column, and so on, then every term in any diagonal array proceeding from upper left to lower right, and consisting of one and only one term from each of the expressions (6), will have the same polynomial coefficient.

I proceed now to obtain expressions for the $L_n$ in (6), in which the individuality of the polynomials $q'_2$, $q'_3$, $q'_4$, etc., has been suppressed. This time I write

$$\log\left(1+re^{at+\frac{bt^2}{2}}\right) = \log\left\{1+r\sum_{k=0}^{\infty}\frac{1}{k!}\left(at+\frac{bt^2}{2}\right)^k\right\}. \tag{8}$$

The term in $t^s$ in $\frac{t^k}{k!}\left(a+\frac{bt}{2}\right)^k$ is

$$\frac{t^s\cdot a^{2k-s}\cdot b^{s-k}}{(2k-s)!(s-k)!\,2^{s-k}}.$$

Rearranging the series in brace of right hand member of (8) in successive powers of $t$,

$$\begin{aligned}\log\left(1+re^{at+\frac{bt^2}{2}}\right) &= \log\left\{1+r\sum_{s=0}^{\infty}\sum_{k=\left[\frac{s+1}{2}\right]}^{s}\frac{a^{2k-s}\,b^{s-k}\cdot t^s}{2^{s-k}(2k-s)!\,(s-k)!}\right\}\\ &= \log(1+r)+\log\left\{1+\theta_1 t+\frac{\theta_2 t^2}{2!}+\cdots+\frac{\theta_s t^s}{s!}+\cdots\right\}.\end{aligned} \tag{9}$$

where $\theta_i = q\beta_i$ $\quad i = 1, 2, 3, \cdots$

Therefore, the coefficient of $\frac{t^s}{s!}$ in the series in the brace in the right member of (9) is $\theta_s$, where

$$\theta_s = q \sum_{k=\left[\frac{s+1}{2}\right]}^{s} \frac{s!\, a^{2k-s}\, b^{s-k}}{2^{s-k}\,(2k-s)!\,(s-k)!}. \tag{10}$$

From equation (2) and (9) above, we have

$$\begin{aligned} &L_1 t + L_2 \frac{t^2}{2!} + L_3 \frac{t^3}{3!} + \cdots \\ &\qquad = \log p + \log(1+r) + \left(m_1 t + \frac{\sigma_1^2 t^2}{2!}\right) + \log\left(1 + \theta_1 t + \frac{\theta_2 t^2}{2!} + \cdots\right). \end{aligned} \tag{11}$$

Therefore, equation (11) becomes

$$L_1 t + L_2 \frac{t^2}{2!} + \cdots = \left(m_1 t + \sigma_1^2 \frac{t^2}{2!}\right) + \log\left(1 + \theta_1 t + \theta_2 \frac{t^2}{2!} + \cdots\right). \tag{12}$$

One might note, in passing, that in (12), the $\theta$'s are playing the rôle of moments, if one recalls the definition of semi-invariants, so that it would be possible to write down a second general expression for the $L$'s, using the well-known formula for semi-invariants in terms of moments.

The first six $\theta$'s take the following form

$$\theta_1 = qa$$

$$\theta_2 = q\,(a^2 + b)$$

$$\theta_3 = q\,(a^3 + 3ab).$$

$$\theta_4 = q\left(a^4 + 6a^2 b + 3b^2\right)$$

$$\theta_5 = q\left(a^5 + 10a^3 b + 15ab^2\right)$$

$$\theta_6 = q\left(a^6 + 15a^4 b + 45a^2 b^2 + 15b^3\right)$$

If, in the last set of relations, we set $q = 1$, we then have $\theta_s = \beta_s$, and if in the expressions (6), $q'_2, q'_3, q'_4$, etc., be all set equal to unity, we shall get, for $n > 2$,

$$L_n = \beta_n .$$

I shall now show that the $\beta$'s follow the recursion law

$$\beta_{s+1} = \left(a + b\frac{\partial}{\partial a}\right)\beta_s . \tag{13}$$

Now, putting $at + \frac{bt^2}{2} = \phi(t)$, we have

$$\frac{d}{dt} e^{\phi(t)} = (a + bt)\cdot e^{\phi(t)} .$$

$$\frac{d^2}{dt^2}\cdot e^{\phi(t)} = \left[b + (a+bt)^2\right]\cdot e^{\phi(t)}$$

$$\frac{d^3}{dt^3}\cdot e^{\phi(t)} = \left[3b(a+bt) + (a+bt)^3\right] e^{\phi(t)} .$$

and in general

$$\frac{d^s}{dt^s}\cdot e^{\phi(t)} = P_s(b, a+bt) . e^{\phi(t)} \tag{14}$$

where $P_s(x,y)$ is a polynomial of degree $s$ in $x$ and $y$.
It is easily shown that

$$e^{\phi(t)}\cdot\frac{d}{dt}\left\{P_s(b, a+bt)\right\}\bigg|_{t=0} = b\frac{\partial}{\partial a}\left\{P_s(b, a+bt)\cdot e^{\phi(t)}\bigg|_{t=0}\right\} . \tag{15}$$

and that

$$(16) \qquad P_s \cdot \frac{d}{dt} \cdot e^{\phi(t)} \Big|_{t=0} = a \cdot P_s \cdot e^{\phi(t)} \Big|_{t=0}$$

Now deriving the left member of (14) with respect to $t$, and then setting $t=0$, gives the next $\beta$, namely $\beta_{s+1}$, by definition, whilst the derivative of the right member of (14), and setting $t=0$, would equal the sum of the right members of (15) and (16), which establishes the proof.

The second property of the expressions for the $L_n$ in (6), which requires proof, may be stated as a theorem thus—"The $k$-th polynomial coefficient in the expression for the semi-invariant $L_{2m}$, $m>1$, is identically equal to the $(k+1)$-st polynomial coefficient in the expression for the semi-invariant $L_{2m+1}$".

For simplicity, I have considered the first and second polynomial coefficients of $L_{2m}$ and $L_{2m+1}$ respectively, the proof going through in exactly the same manner if perfectly general terms in these expressions were considered.

From (5), suppose that $n=2m$ (even). Then

$$(17) \quad L_{2m} = \sum_{k=1}^{2m} \sum_{i=1}^{k} \frac{(-1)^{1+i} \cdot q^k}{k} \binom{k}{i} \sum_{j=m}^{2m} \frac{(2m)!\, a^{2j-2m} \cdot b^{2m-j} \cdot i^j}{2^{2m-j} \cdot (2j-2m)!\,(2m-j)!}$$

The leading term in $L_{2m}$, i.e. the first term in $\beta_{2m}$, multiplied by a polynomial in $q$, is obtained from (17), by setting $j=2m$. It is

$$\sum_{k=1}^{2m} \sum_{i=1}^{k} \frac{(-1)^{1+i} \cdot q^k}{k} \binom{k}{i} a^{2m} \cdot i^{2m} .$$

Therefore the polynomial coefficient of $a^{2m}$ i.e. the leading coefficient in $L_{2m}$ is

$$(18) \qquad \sum_{k=1}^{2m} \sum_{\iota=1}^{k} \frac{(-1)^{1+\iota} \cdot q^k}{k} \cdot \binom{k}{\iota} \cdot \iota^{2m}$$

Again, from (5), when $n = 2m+1$ (odd),

$$L_{2m+1} = \sum_{k=1}^{2m+1} \sum_{\iota=1}^{k} \frac{(-1)^{1+\iota} \cdot q^k}{k} \binom{k}{\iota} \sum_{j=m+1}^{2m+1} \frac{(2m+1)!\, a^{2j-(2m+1)} \cdot b^{(2m+1)-j} \cdot \iota^j}{2^{(2m+1)-j} \cdot [2j-(2m+1)]!\,[2m+1-j]!} .$$

The second term in $L_{2m+1}$ is obtained by setting $j = 2m$. It is

$$\sum_{k=1}^{2m+1} \sum_{i=1}^{k} \frac{(-1)^{1+\iota} \cdot q^k}{k} \cdot \binom{k}{i} \cdot m(2m+1) \cdot a^{2m-1} \cdot b \cdot \iota^{2m} .$$

Therefore, the polynomial coefficient of $a^{2m-1} \cdot b$ i.e. the second coefficient in $L_{2m+1}$ is

$$(18') \qquad \sum_{k=1}^{2m+1} \sum_{i=1}^{k} \frac{(-1)^{1+\iota} \cdot q^k}{k} \cdot \binom{k}{i} \cdot \iota^{2m} .$$

Comparing (18) and (18′), which must be identically equal, if our theorem is true, it remains to show that

$$\frac{q^{2m+1}}{2m+1} \cdot \sum_{i=1}^{2m+1} (-1)^{1+\iota} \cdot \binom{2m+1}{i} \, \iota^{2m} = 0 .$$

but it is well known that this expression is identically zero [2]

[2] See Hall and Knight. Higher Algebra, p 259, Ex. 2.

*Section 2. A table of values of a certain class of polynomials in one variable, for different values of the argument.*

In order to facilitate the actual computation of the values for the semi-invariants $L_n$, given in Section 1, in a particular application of the theory, when $a = m_2 - m_1$, and $b = \sigma_2^2 - \sigma_1^2$, are known, I consider the expressions for the $L_n$, as they appear in the form indicated in (6). Now, when $b = \sigma_2^2 - \sigma_1^2 = 0$, i.e. the two components have identical standard deviations, the set of relations (6) take the form

$$(19) \quad L'_1 = m_1 + aq \cdot L'_2 = \sigma_1^2 + a^2 pq \text{ and } L'_n = a^n pq q_n \;;\; n > 2$$

in which $q_3$, $q_4$, etc., have the same significance as in (7). Making use of the properties of the expressions (6), which were stated at the end of Section 1, from (19) we may write the $L_n$, for $n > 4$, as follows

$$L_5 = L'_5 + 10\left(\frac{b}{a}\right)L'_4 + 15\left(\frac{b}{a}\right)^2 L'_3 .$$

$$L_6 = L'_6 + 15\left(\frac{b}{a}\right)L'_5 + 45\left(\frac{b}{a}\right)^2 L'_4 + 15\left(\frac{b}{a}\right)^3 L'_3 .$$

$$(20) \quad L_7 = L'_7 + 21\left(\frac{b}{a}\right)L'_6 + 105\left(\frac{b}{a}\right)^2 L'_5 + 105\left(\frac{b}{a}\right)^3 L'_4 .$$

$$L_8 = L'_8 + 28\left(\frac{b}{a}\right)L'_7 + 210\left(\frac{b}{a}\right)^2 L'_6 + 420\left(\frac{b}{a}\right)^3 L'_5 + 105\left(\frac{b}{a}\right)^4 L'_4 .$$

$$L_9 = L'_9 + 36\left(\frac{b}{a}\right)L'_8 + 378\left(\frac{b}{a}\right)^2 L'_7 + 1260\left(\frac{b}{a}\right)^3 L'_6 + 945\left(\frac{b}{a}\right)^4 L'_5 .$$

and so on.

Therefore, for $n > 4$, the general semi-invariants $L_n$ of (6)

may be expressed in terms of the special semi-invariants $L'_n$, obtained from the former by setting $b=0$. From (20) in general we have

$$L_n = \sum_{k=0}^{n} L'_{n-k} \quad \frac{n^{(2k)}}{2^k \cdot k!} \cdot \left(\frac{b}{a}\right)^k,$$

in which

$$L'_{n-k} = a^{n-k} pq q_{n-k} \,,$$

because

$$L_n = \sum_{k=0}^{\left[\frac{n-1}{2}\right]} a^{n-k} pq q_{n-k} \cdot \frac{n^{(2k)}}{2^k \cdot k!} \left(\frac{b}{a}\right)^k$$

$$= \sum_{k=0}^{\left[\frac{n-1}{2}\right]} a^{n-2k} \cdot b^k \cdot \frac{n^{(2k)}}{2^k \cdot k!} \, q_{n-k} \,,$$

and the last expression, for $n > 4$, is the equivalent of the general expression (5) for the $L_n$.

Further, if we consider the terms in the expressions (20) as elements in a set of diagonal arrays, as I have indicated, it is evident that, moving downwards along any particular diagonal, any term in this diagonal is obtained from the one immediately preceding it, by the use of a multiplier $M_{k\ell} \cdot \left(\frac{b}{a}\right)$. The formula for the calculation of the $M_{k\ell}$, may be derived as follows.

Consider any term of say $L_n$, whose numerical coefficient is $C_{k+1,n}$. Let this be the $(k+1)$st term.

Then

$$C_{k+1,n} = \frac{n^{(2k)}}{2^k \cdot k!}$$

Similarly, take the $(k+2)$nd term of $L_{n+1}$, with numerical coefficient $C_{k+2,n+1}$. Then

$$C_{k+2,n+1} = \frac{(n+1)^{(2k+2)}}{2^{k+1}\cdot(k+1)!},$$

and we note that $C_{k+1,n}$ and $C_{k+2,n+1}$ are the numerical coefficients of two adjacent terms in one of the diagonal arrays mentioned above. Therefore

$$C_{k+2,n+1} = \frac{(n+1)(n-2k)}{2(k+1)}\cdot C_{k+1,n} \qquad n \geqq 2k$$

and

$$M_{k+1,n} = \frac{(n+1)(n-2k)}{2(k+1)}.$$

It is of considerable interest to note, that the $L'_n$ of (19) (for $n > 2$), are, except perhaps for sign, the product of the semi-invariants $\lambda_n$ $(n > 1)$ of the binomial distribution and appropriate powers of "a". To show this, we need only consider the generating function for the $L'_n$, namely $1+q(e^{at}-1)$, and that for the $\lambda_n$, viz $\left[1+p(e^t-1)\right]^s$, with $s=1$. Frisch[1] has obtained a recursion formula for the $\lambda_n$, which is

$$\lambda_n = pq\cdot\frac{d}{dp}\cdot\lambda_{n-1}, \qquad n>1$$

[1] Sur les semi-invariants et moments employés dans l'étude des distributions statistiques. Ragnar Frisch. Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo. 1926. No. 3, Ch. 2, p. 29

Therefore, it is evident that the $L'_n$ obey a corresponding recursion formula

$$L'_n = a^n \; pq \cdot \frac{d}{dq} \; L'_{n-1} \qquad n > 2$$

In fact, the polynomials $q'_2 = pq$, $q'_3 = pqq_3$, $q'_4 = pqq_4$, etc., are the same functions of $q$ as the $\lambda_n$ are functions of p, i e.

$$\lambda_n = \lambda_n(p). \;, \; q'_n = \lambda_n(q). \qquad \text{for } n \geq 2.$$

To investigate thoroughly the properties of the polynomials[2] $q'_2$, $q'_3$, $q'_4$, etc , would be irrelevant to the problem in hand, but, so far as I know, such a study has not been carried out. I will, however, mention a few of these properties, which appear interesting.

1. The roots of the polynomial $q'_n$ (for any $n \geq 2$ ) are all real and distinct, and these roots all lie in the interval (0, 1), zero and unity being roots of every polynomial.
2. The roots of $q'_n$ separate the roots of $q'_{n+1}$
3. The polynomials $q'_{2n}$, of even degree in $q$ , are symmetrical with respect to the line $x = \frac{1}{2}$ , whilst those of odd degree in $q$ , namely $q'_{2n+1}$ , are symmetrical with respect to the point (½, 0)
4. An orthogonality property in (0, 1) holds if $m \neq n$, and $m+n$ is odd. That is

$$\int_0^1 q'_m q'_n = 0. \qquad m \neq n, \; m+n \text{(odd)}$$

but $$\int_0^1 q'_m q'_n \neq 0. \qquad m \neq n, \; m+n \text{(even)}$$

[2] These same polynomials appear as functions of $p$ in a paper by H C. Carver on "Fundamentals of the theory of Sampling" Amer Statist Assoc. Annals of Math. Statistics. Vol I, No. 1, Feb. 1930, p. 106.

5. Further

$$\int_0^1 q'_{2n} \cdot dq = (-1)^{n-1} \cdot B_{2n}$$

in which $B_{2n}$ is the Bernoulli number of order $2n$.

$$\int_0^1 q'_{2n+1} \cdot dq = 0.$$

In calculating the actual values of the $L_n$ expressed in the form (20), it would obviously be very convenient to have at one's disposal a table of values of the polynomials $q'_2 = pq$, $q'_3 = pqq_3$, $q'_4 = pqq_4$, etc., for a range of values of $q$, since the latter, when multiplied by appropriate powers of "a", are the $L'_n$ of (19) and (20). I have, therefore, set up such tables, for values of the variable $q$ ranging from $.01$ to $1.0$ inclusive, at intervals of $.01$.

It is to be observed that only functional values are recorded here for $.01 \leqq q \leqq .50$, since we would merely repeat these values when $.50 \leqq q \leqq 1.0$, in the case of the polynomials $q'_{2n}$, of even degree, whilst in the case of those of odd degree, namely $q'_{2n+1}$, there would merely be a change of sign. For it is easily seen that, writing, $q'_n = \lambda_n(q)$,

$$\lambda_n(q) = (-1)^n \lambda_n(p), \qquad (p+q) = 1.$$

Hence

$$\lambda_{2n}(q) = \lambda_{2n}(p)$$

and

$$\lambda_{2n+1}(q) = -\lambda_{2n+1}(p).$$

I have calculated the exact functional values of all the polynomials $q'_n$, for $2 \leqq n \leqq 9$ and these values appear in the tables

for those cases in which $n \leqq 4$, but for $n > 4$ the functional values are written down correct only to eight decimal places Each polynomial is set out in detail below.

$$q'_2 = q - q^2 :$$

$$q'_3 = q - 3q^2 + 2q^3.$$

$$q'_4 = q - 7q^2 + 12q^3 - 6q^4.$$

$$q'_5 = q - 15q^2 + 50q^3 - 60q^4 + 24q^5.$$

$$q'_6 = q - 31q^2 + 180q^3 - 390q^4 + 360q^5 - 120q^6.$$

$$q'_7 = q - 63q^2 + 602q^3 - 2100q^4 + 3360q^5 - 2520q^6 + 720q^7.$$

$$q'_8 = q - 127q^2 + 1932q^3 - 10206q^4 + 25200q^5 - 31920q^6$$

$$+ 20160q^7 - 5040q^8.$$

$$q'_9 = q - 255q^2 + 6050q^3 - 46620q^4 + 166824q^5$$

$$- 317520q^6 + 332640q^7 - 181440q^8 + 40320q^9$$

| $q$ | $q'_2$ | $q'_3$ | $q'_4$ | $q'_5$ |
|---|---|---|---|---|
| .01 | 0099 | .0097 02 | .0093 1194 | 0085 4940 |
| .02 | 0196 | .0188 16 | .0172 9504 | .0143 9048 |
| .03 | .0291 | 0273 54 | 0240 1914 | .0178 0198 |
| .04 | .0384 | 0353 28 | .0295 5264 | .0190 4886 |
| .05 | .0475 | 0427 50 | 0339 6250 | .0183 8250 |
| .06 | .0564 | 0496 32 | 0373 1424 | 0160 4106 |
| .07 | .0651 | .0559 86 | 0396 7194 | .0122 4974 |
| .08 | .0736 | 0618 24 | 0410 9824 | .0072 2104 |
| .09 | .0819 | 0671 58 | .0416 5434 | .0011 5512 |
| .10 | .0900 | .0720 00 | 0414 0000 | −.0057 6000 |
| .11 | .0979 | .0763 62 | 0403 9354 | − 0133 4808 |
| .12 | 1056 | .0802 56 | 0386 9184 | − 0214 4440 |
| .13 | 1131 | 0836 94 | 0363 5034 | −.0298 9550 |
| .14 | .1204 | 0866 88 | .0334 2304 | −.0385 5882 |
| .15 | .1275 | 0892 50 | 0299 6250 | −.0473 0250 |
| .16 | 1344 | .0913 92 | 0260 1984 | −.0560 0502 |
| .17 | 1411 | .0931 26 | .0216 4474 | −.0645 5494 |
| .18 | .1476 | 0944 64 | 0168 8544 | −.0728 5064 |
| .19 | .1539 | 0954 18 | 0117 8874 | −.0807 9996 |
| .20 | .1600 | .0960 00 | .0064 0000 | −.0883 2000 |
| .21 | .1659 | .0962 22 | 0007 6314 | −.0953 3676 |
| .22 | 1716 | .0960 96 | −.0050 7936 | −.1017 8488 |
| .23 | .1771 | .0956 34 | −.0110 8646 | −.1076 0738 |
| 24 | 1824 | .0948 48 | −.0172 1856 | −.1127 5530 |
| .25 | .1875 | 0937 50 | −.0234 3750 | −.1171 8750 |
| .26 | 1924 | 0923 52 | −.0297 0656 | −.1208 7030 |
| .27 | .1971 | .0906 66 | −.0359 9046 | −.1237 7722 |
| .28 | 2016 | .0887 04 | −.0422 5536 | −.1258 8872 |
| .29 | .2059 | 0864 78 | −.0484 6886 | −.1271 9184 |
| 30 | .2100 | .0840 00 | −.0546 0000 | −.1276 8000 |
| .31 | 2139 | 0812 82 | − 0606 1926 | −.1273 5264 |
| .32 | .2176 | .0783 36 | − 0664 9856 | − 1262 1496 |
| .33 | .2211 | .0751 74 | − 0722 1126 | −.1242 7766 |
| .34 | .2244 | .0718 08 | −.0777 3216 | − 1215 5658 |
| .35 | 2275 | .0682 50 | −.0830 3750 | − 1180 7250 |
| .36 | 2304 | 0645 12 | −.0881 0496 | −.1138 5078 |
| .37 | 2331 | .0606 06 | −.0929 1366 | −.1089 2110 |
| 38 | .2356 | .0565 44 | −.0974 4416 | −.1033 1720 |
| .39 | .2379 | 0523 38 | − 1016 7846 | −.0970 7652 |
| .40 | .2400 | 0480 00 | −.1056 0000 | −.0902 4000 |
| .41 | .2419 | .0435 42 | − 1091 9366 | −.0828 5172 |
| .42 | .2436 | 0389 76 | −.1124 4576 | −.0749 5864 |
| .43 | .2451 | 0343 14 | − 1153 4406 | −.0666 1034 |
| .44 | .2464 | 0295 68 | − 1178 7776 | −.0578 5866 |
| .45 | .2475 | .0247 50 | −.1200 3750 | −.0487 5750 |
| .46 | 2484 | .0198 72 | −.1218 1536 | − 0393 6246 |
| .47 | .2491 | .0149 46 | − 1232 0486 | − 0297 3058 |
| .48 | .2496 | .0099 84 | − 1242 0096 | −.0199 2008 |
| .49 | .2499 | .0049 98 | −.1248 0006 | −.0099 9000 |
| .50 | .2500 | .0000 00 | −.1250 0000 | .0000 0000 |

*Section 3. Approximate expressions for the semi-invariants of* $\alpha_3$, $\alpha_4$, $\sigma_x$ *in samples from the compound frequency function.*

In the paper of C. C. Craig, already cited, the author obtained the following results for sampling from a single parent population.

(1) Expressions[1] for the sampling characteristics of the correlation functions for $\nu_2, \nu_3$, and $\nu_2$, $\nu_4$, in terms of $N$, the size of the sample, and the characteristics of the population itself.

(2) Expressions[2] for the sampling characteristics of the distribution functions for $\alpha_3$, $\alpha_4$, and $\sigma_x$, in terms of certain "g" functions, the latter being defined by

$$g_{k\ell}(\nu_m, \nu_n) = \frac{S_{k\ell}(\nu_m, \nu_n)}{\nu_2^{\frac{km+\ell n}{2}}} \tag{21}$$

in which $S_{k\ell}(\nu_m, \nu_n)$ are the characteristics of the correlation function for $\nu_m$, $\nu_n$.

I can now make use of the results indicated above, in conjunction with the relations (6) of the present paper, in order to determine approximate expressions for the semi-invariants of $\alpha_3$, $\alpha_4$, $\sigma_x$, in samples from the compound frequency function, retaining only terms of order $-2$ and higher in $N$ in using expressions (1), and only those "g" functions in using the expressions (2) which are of order $-2$ and higher in $N$, where $g_{k\ell}$ is of order $\frac{1}{N^{k+\ell-1}}$.

A. The semi-invariants of $\alpha_3$, viz. $b_1, b_2, b_3$, etc.

From definition (21) and the relations (1) for $m=2$, $n=3$, and making use of the following notation

[1] Loc. Cit p. 57 et seq.
[2] Loc. Cit. p. 50 et seq.

$$(22) \qquad \phi_i = \frac{L_i}{L_2^{\frac{i}{2}}}, \qquad \phi_{ikj}^{rst} = \phi_i^r \cdot \phi_k^s \; \phi_j^t, \qquad \text{etc.}$$

in which the $L_n$ are the same as in (6) of Section 1, I obtain the following set of "g" functions

using for brevity $g_{k\ell} = g_{k\ell}(v_2, v_3)$.

$$(23) \quad \begin{aligned}
g_{10} &= \frac{1}{N}\left\{(N-1)\right\} \\
g_{01} &= \frac{1}{N^2}\left\{(N-1)(N-2)\phi_3\right\} \\
g_{20} &= \frac{1}{N^3}\left\{(N-1)^2\phi_4 + 2N(N-1)\right\} \\
g_{11} &= \frac{1}{N^4}\left\{(N-1)^2(N-2)\phi_5 + 6N(N-1)(N-2)\phi_3\right\} \\
g_{02} &= \frac{1}{N^5}\left\{(N-1)^2(N-2)^2\phi_6 + 9N(N-1)(N-2)^2\phi_4 + \right. \\
&\qquad \left. 9N(N-1)(N-2)^2\phi_3^2 + 6N^2(N-1)(N-2)\right\}. \\
g_{30} &= \frac{1}{N^5}\left\{(N-1)^3\phi_6 + 12N(N-1)^2\phi_4 + 4N(N-1)(N-2)\phi_3^2 + 8N^2(N-1)\right\} \\
g_{21} &= \frac{1}{N^6}\left\{(N-1)^3(N-2)\phi_7 + 16N(N-1)^2(N-2)\phi_5 \right. \\
&\qquad \left. + 12N(N-1)(N-2)(2N-3)\phi_{43} + 48N^2(N-1)(N-2)\phi_3\right\}. \\
g_{12} &= \frac{1}{N^7}\left\{(N-1)^3(N-2)^2\phi_8 + 21N(N-1)^2(N-2)^2\phi_6 \right. \\
&\qquad + 6N(N-1)(N-2)^2(8N-11)\phi_{53} + 9N(N-1)(N-2)^2(3N-5)\phi_4^2 \\
&\qquad + 18N^2(N-1)(N-2)(6N-11)\phi_4 + 18N^2(N-1)(N-2)(9N-20)\phi_3^2 \\
&\qquad \left. + 36N^3(N-1)(N-2)\right\}
\end{aligned}$$

$$g_{03}=\frac{1}{N^8}\Big\{(N-1)^3(N-2)^3\phi_9+27N(N-1)^2(N-2)^3\phi_7+27N(N-1)(N-2)^3(3N-4)\phi_{63}$$

$$+27N(N-1)(N-2)^3(4N-7)\phi_{54}+54N^2(N-1)(N-2)^2(4N-7)\phi_5$$

$$+162N^2(N-1)(N-2)^2(5N-12)\phi_{43}+36N^2(N-1)(N-2)(7N^2-30N+34)\phi_3^3$$

$$+108N^3(N-1)(N-2)(5N-12)\phi_3\Big\}.$$

and on substituting these values for the "g"s in the expressions for the semi-invariants of $\alpha_3$ from the relations (2), I get :—

$$b_1=\alpha_3\Big\{\frac{13}{16}+\frac{1}{N}\Big(\frac{783}{32}+\frac{645}{64}\phi_4\Big)+\frac{1}{N^2}\Big(-\frac{9155}{64}-\frac{5385}{64}\phi_4-\frac{8505}{256}\phi_4^2+\frac{245}{32}\phi_6$$

$$+\frac{245}{8}\phi_3^2\Big)\Big\}+\Big\{\frac{11}{8}\phi_3+\frac{1}{N}\Big(\frac{-333}{8}\phi_3-\frac{69}{16}\phi_5-\frac{75}{16}\phi_{34}\Big)+\frac{1}{N^2}\Big(\frac{8273}{32}\phi_3$$

$$-\frac{39}{16}\phi_5-\frac{75}{16}\phi_7+\frac{1875}{32}\phi_{34}-\frac{35}{16}\phi_{36}-\frac{35}{4}\phi_3^2+\frac{735}{32}\phi_{45}+\frac{945}{128}\phi_{34}^{12}\Big)\Big\}.$$

$$b_2=\alpha_3^2\Big\{\frac{1}{N}\Big(\frac{99}{2}+\frac{99}{4}\phi_4\Big)+\frac{1}{N^2}\Big(-\frac{2313}{4}-\frac{819}{2}\phi_4+\frac{225}{8}\phi_6+\frac{225}{2}\phi_3^2-\frac{2565}{16}\phi_4^2\Big)\Big\}$$

$$+2\alpha_3\Big\{\frac{1}{N}\Big(-81\phi_3-\frac{21}{2}\phi_5-9\phi_{34}\Big)+\frac{1}{N^2}\Big(\frac{7479}{8}\phi_3+\frac{33}{4}\phi_5-\frac{33}{2}\phi_7+\frac{2547}{8}\phi_{34}$$

$$-\frac{45}{8}\phi_{36}-\frac{45}{2}\phi_3^3+\frac{825}{8}\phi_{45}+\frac{855}{32}\phi_{34}^{12}\Big)\Big\}+\Big\{\frac{1}{N}\Big(24+36\phi_4+4\phi_6+9\phi_{35}$$

$$+\frac{189}{2}\phi_3^2+\frac{9}{4}\phi_{34}^{21}\Big)+\frac{1}{N^2}\Big(-90+135\phi_4+108\phi_6+9\phi_8+27\phi_4^2-\frac{2331}{2}\phi_3^2$$

$$-39\phi_5^2-\frac{567}{2}\phi_{34}^{21}-\frac{123}{2}\phi_{35}+\frac{33}{4}\phi_{37}-\frac{165}{4}\phi_{345}-24\phi_{46}\Big)\Big\}.$$

$$b_3=\alpha_3^3\Big\{\frac{1}{N^2}\Big(\frac{4671}{8}-\frac{3969}{8}\phi_4+\frac{351}{16}\phi_6+\frac{351}{4}\phi_3^2-\frac{6075}{32}\phi_4^2\Big)\Big\}$$

$$+3\alpha_3^2\Big\{\frac{1}{N^2}\Big(\frac{3537}{4}\phi_3+36\phi_5-\frac{99}{8}\phi_7+\frac{1863}{4}\phi_{34}-\frac{27}{8}\phi_{36}-\frac{27}{2}\phi_3^3+117\phi_{45}\Big)\Big\}$$

$$+3\alpha_3\Big\{\frac{1}{N^2}\Big(-54+135\phi_4+\frac{369}{4}\phi_6-\frac{2403}{2}\phi_3^2+\frac{27}{4}\phi_8-\frac{441}{2}\phi_{35}-\frac{81}{2}\phi_4^2$$

$$+\frac{9}{2}\phi_{37}-\frac{1323}{4}\phi_{34}^{21}-36\phi_{345}-\frac{99}{4}\phi_{46}-\frac{363}{8}\phi_5^2\Big)\Big\}+\Big\{\frac{1}{N^2}\Big(-432\phi_3$$

$$-513\phi_5-\frac{189}{2}\phi_7-810\phi_{34}-\frac{7}{2}\phi_9-108\phi_{36}-\frac{27}{2}\phi_{45}+1710\phi_3^3$$

$$-\frac{9}{2}\phi_{38}+\frac{891}{2}\phi_{35}^{21}+\frac{81}{2}\phi_{56}+\frac{27}{2}\phi_{346}+\frac{243}{2}\phi_{34}^{31}+\frac{99}{4}\phi_{35}^{12}\Big)\Big\}$$

$b_4=0$

In a similar manner I obtain

B. The semi-invariants of $\alpha_4$, viz $c_1$, $c_2$, $c_3$, etc.

In this case $m=2$, $n=4$, and the "g" functions this time are

$$g_{10}=\frac{1}{N}\Big\{(N-1)\Big\}. \qquad \text{for brevity } g_{k\ell}\equiv g_{k\ell}(v_2,v_4)$$

$$g_{01}=\frac{1}{N^3}\Big\{N(N^2-4N+6)\phi_4+3N(N^2-2N+1)\Big\}.$$

$$g_{20}=\frac{1}{N^3}\Big\{N(N-2)\phi_4+2N(N-1)\Big\}$$

$$g_{11}=\frac{1}{N^2}\Big\{(N-5)\phi_6+2(7N-25)\phi_4+6(N-5)\phi_3^2+12(N-2)\Big\}$$

$$g_{02}=\frac{1}{N^2}\Big\{(N-8)\phi_8+4(7N-46)\phi_6+48(N-8)\phi_{35}+2(17N-128)\phi_4^2 \tag{24}$$

$$+12(17N-88)\phi_4+72(3N-20)\phi_3^2+24(4N-13)\Big\}$$

$$g_{30}=\frac{1}{N^2}\Big\{\phi_6+12\phi_4+4\phi_3^2+8\Big\}$$

$$g_{21}=\frac{1}{N^2}\Big\{\phi_8+26\phi_6+40\phi_{35}+34\phi_4^2+176\phi_4+144\phi_3^2+72\Big\}$$

$$g_{12}=\frac{1}{N^2}\Big\{\phi_X+44\phi_{28}+56\phi_{37}+208\phi_{46}+116\phi_5^2+596\phi_{26}^{21}$$

$$+2016\,\phi_{235}+1512\,\phi_{24}^{12}+1584\,\phi_{34}^{21}+2832\,\phi_{24}^{31}+4176\,\phi_{23}^{22}+768\,\phi_{2}^{5}\Big\}$$

$$g_{03}=\frac{1}{N^2}\Big\{\phi_{\underline{\mathrm{XII}}}+66\,\phi_{2\underline{\mathrm{X}}}+208\,\phi_{39}+492\,\phi_{48}+768\,\phi_{57}+462\,\phi_{6}^{2}+1476\,\phi_{28}^{21}$$

$$+7344\,\phi_{237}+13776\,\phi_{246}+7848\,\phi_{25}^{12}+8280\,\phi_{36}^{21}+25344\,\phi_{345}$$

$$+5672\,\phi_{4}^{3}+11448\,\phi_{26}^{31}+73440\,\phi_{235}^{211}+51048\,\phi_{24}^{22}+119232\,\phi_{234}^{121}$$

$$+12456\,\phi_{3}^{4}+49248\,\phi_{24}^{41}+110592\,\phi_{23}^{32}+9504\,\phi_{2}^{6}\Big\}$$

Therefore, substituting from (24) into the expressions for the semi-invariants of $\alpha_4$ given by the relations (2) we have—

$$c_1=\alpha_4\Big\{-2+\frac{1}{N}\left(50+21\phi_4\right)+\frac{1}{N^2}\left(401+16\phi_6-75\phi_4^2+522\,\phi_4+64\,\phi_3^2\right)\Big\}$$

$$+\Big\{\left(6+2\phi_4\right)+\frac{1}{N}\left(-174-8\phi_6-9\phi_4^2-169\phi_4-48\phi_3^2\right)+\frac{1}{N^2}\Big(1269-9\,\phi_8$$

$$-92\,\phi_6+489\phi_4^2+15\phi_4^3-420\phi_3^2-360\phi_{35}+272\,\phi_{34}^{21}+1341\,\phi_4+44\,\phi_{46}\Big)\Big\}.$$

$$c_2=\alpha_4^2\Big\{\frac{1}{N}\left(128+64\,\phi_4\right)+\frac{1}{N^2}\left(-1688+72\,\phi_6-462\,\phi_4^2-1256\,\phi_4+288\,\phi_3^2\right)\Big\}$$

$$+2\alpha_4\Big\{\frac{1}{N}\left(-384-22\phi_6-20\phi_4^2-408\,\phi_4-132\,\phi_3^2\right)+\frac{1}{N^2}\Big(5664-35\,\phi_8$$

$$-282\,\phi_6+2800\,\phi_4^2+6808\,\phi_4-1200\,\phi_3^2-1400\,\phi_{35}+1464\,\phi_{34}^{21}$$

$$+240\,\phi_{46}+66\,\phi_4^3\Big)\Big\}+\Big\{\frac{1}{N}\Big(1320+7\phi_8+244\,\phi_6+494\,\phi_4^2+4\,\phi_4^3+2376\,\phi_4$$

$$+1800\,\phi_3^2+336\,\phi_{35}+16\,\phi_{46}+96\,\phi_{34}^{21}\Big)+\frac{1}{N^2}\Big(-32208-140\,\phi_8-6544\,\phi_6$$

$$-2440\,\phi_4^3-37192\,\phi_4^2-75000\,\phi_4-48168\,\phi_3^2-7056\,\phi_{35}+704\,\phi_{28}$$

$$+896\,\phi_{37}-532\,\phi_{46}+1856\,\phi_5^2+9536\,\phi_{26}^{21}+32256\,\phi_{235}+24192\,\phi_{24}^{12}$$

$$-384\,\phi_{34}^{21}+16\,\phi_{\underline{\mathrm{X}}}+45312\,\phi_{24}^{31}+66816\,\phi_{23}^{22}+12288\,\phi_2^5-36\,\phi_{48}-1840\,\phi_{345}$$

$$
\begin{aligned}
&-80\phi_6^2-2880\phi_3^4+11920\phi_{26}^{21}-960\phi_{36}^{21}-84\phi_{46}^{21}-504\phi_{34}^{22}\Big)\Big\}.\\
&c_3=\frac{1}{N^2}\Big\{\alpha_4^3\left(-2080+64\phi_6-648\phi_4^2-1824\phi_4+256\phi_3^2\right)\\
(25)\quad &+3\alpha_4^2\Big(7488-28\phi_8-16\phi_6+4608\phi_4^2+10592\phi_4+288\phi_3^2\\
&-1120\phi_{35}+360\phi_{46}+2176\phi_{34}^{21}\Big)+3\alpha_4\Big(-28896+12\phi_{\underline{\mathrm{X}}}-88\phi_8\\
&-5344\phi_6-2752\phi_4^3-38776\phi_4^2-67872\phi_4-37728\phi_3^2-3528\phi_3^4\\
&+528\phi_{28}+672\phi_{37}-2008\phi_{46}+1392\phi_5^2+9216\phi_2^5+7152\phi_{26}^{21}\\
&+24192\phi_{235}+18144\phi_{24}^{12}-10800\phi_{34}^{21}+33984\phi_{24}^{31}+50112\phi_{23}^{22}\\
&-4416\phi_{35}+7152\phi_{26}^{21}-48\phi_{48}-2368\phi_{345}-98\phi_6^2-1176\phi_{36}^{21}\\
&-80\phi_{46}^{21}-480\phi_{34}^{22}\Big)+\Big(114912-5\phi_{\overline{\underline{\mathrm{XII}}}}-18\phi_{\mathrm{X}}-330\phi_{2\mathrm{X}}-6\phi_{4\mathrm{X}}-1040\phi_{39}\\
&-1332\phi_{48}-3840\phi_{57}-168\phi_6^2-7380\phi_{28}^{21}-36720\phi_{237}-68880\phi_{246}\\
&-39240\phi_{25}^{12}-12240\phi_{36}^{21}-72576\phi_{345}+23120\phi_4^3-57240\phi_{26}^{31}\\
&-367200\phi_{235}^{211}-255240\phi_{24}^{22}-596160\phi_{234}^{121}+35568\phi_3^4-246240\phi_{24}^{41}\\
&-552960\phi_{23}^{32}-47520\phi_2^6-792\phi_{28}-1008\phi_{37}+47064\phi_{46}-2088\phi_5^2\\
&-10728\phi_{26}^{21}-36288\phi_{235}-27216\phi_{24}^{12}+330480\phi_{34}^{21}-50976\phi_{24}^{31}\\
&-75168\phi_{23}^{22}-13824\phi_2^5-264\phi_{248}-336\phi_{347}+3048\phi_{46}^{21}-696\phi_{45}^{12}\\
&-3576\phi_{246}^{211}-12096\phi_{2345}-9072\phi_{24}^{13}+17424\phi_{34}^{22}-16992\phi_{24}^{32}\\
&-25056\phi_{234}^{221}-4608\phi_{24}^{51}+72\phi_{68}+432\phi_{38}^{21}+1008\phi_8+38160\phi_6\\
&+3456\phi_{356}+20736\phi_{35}^{31}+48384\phi_{35}+305496\phi_4^2+362304\phi_4\\
&+277344\phi_3^2+24\phi_{48}^{21}+1152\phi_{345}^{121}+42\phi_{46}^{12}+1512\phi_{34}^{41}+504\phi_{346}^{211}+816\phi_4^4\Big)\Big\}.
\end{aligned}
$$

C. The semi-invariants of $\sigma_x (=\sqrt{V_2})$, viz. $d_1$, $d_2$, $d_3$, etc.

Now $g_r \equiv g_{r0}$ and $g_r(V_2) = \dfrac{S_r(V_2)}{L_2^r}$ from the relation (21)

Therefore, the "g" functions here are—

$$g_1 = \frac{1}{N}\left\{(N-1)\right\}, \qquad \text{for brevity } g_r = g_r(V_2)$$

$$g_2 = \frac{1}{N^2}\left\{(N-2)\phi_4 + 2(N-1)\right\}, \tag{26}$$

$$g_3 = \frac{1}{N^2}\left\{\phi_6 + 12\phi_4 + 4\phi_3^2 + 8\right\}.$$

On substituting from (26) into the expressions for the semi-invariants of $\sigma_x$ from the relations (2) gives—

$$d_1 = \sigma_x\left\{\frac{23}{16} + \frac{1}{N}\left(-\frac{25}{32} - \frac{11}{64}\phi_4\right) + \frac{1}{N^2}\left(\frac{89}{64} - \frac{3}{32}\phi_6 + \frac{43}{64}\phi_4 - \frac{3}{8}\phi_3^2 + \frac{75}{256}\phi_4^2\right)\right\},$$

$$d_2 = \frac{L_2}{4}\left\{\frac{1}{N}\left(2+\phi_4\right) + \frac{1}{N^2}\left(-7 + \frac{1}{2}\phi_6 - 4\phi_4 + 2\phi_3^2 - \frac{7}{4}\phi_4^2\right)\right\},$$

$$d_3 = \frac{\sigma_x^3}{8}\left\{\frac{1}{N^2}\left(5 - \frac{1}{2}\phi_6 + 3\phi_4 - 2\phi_3^2 + \frac{9}{4}\phi_4^2\right)\right\},$$

$$d_4 = 0.$$

*Section 4. The case in which the compound frequency function may possess non-zero semi-invariants of all orders.*

Instead of the components of the compound frequency function being considered normal, I now assume that they may possess non-zero semi-invariants as far as the third order, and I again derive expressions for the parameters of the compound in terms of the parameters of the components. The method of derivation is entirely analogous to that in Section 1, where the components were normal. The expressions for the $L_n$, the semi-invariants of the compound are seen to be more complicated in the present case.

but this complexity is more apparent than real. In fact, I have succeeded in deducing a rather simple general law, by means of which, these expressions may be written out, and this law is still applicable, even if the two components should possess non-zero semi-invariants of higher orders than the third.

I now write

$$(27)\quad e^{L_1 t + L_2\frac{t^2}{2} + L_3\frac{t^3}{3!}+\cdots} = p e^{m_1 t + \sigma_1^2\frac{t^2}{2} + {}_1\lambda_3\frac{t^3}{3!}}\left\{1 + r e^{at + \frac{bt^2}{2!} + \frac{ct^3}{3!}}\right\},$$

where

$$a = m_2 - m_1, \qquad b = \sigma_2^2 - \sigma_1^2, \qquad c = {}_2\lambda_3 - {}_1\lambda_3, \qquad r = \frac{q}{p},$$

and $m_1, m_2, \sigma_1, \sigma_2$, have the same significance as in Section 1, and ${}_1\lambda_3, {}_2\lambda_3$, are the third semi-invariants of $\phi_1(x)$, $\phi_2(x)$ respectively, and as before $p+q=1$. Taking logarithms in (27), we have

$$(28)\quad \begin{aligned} &L_1 t + L_2\frac{t^2}{2} + L_3\frac{t^3}{3!}+\cdots \\ &= \log p + \left(m_1 t + \frac{\sigma_1^2 t^2}{2} + \frac{{}_1\lambda_3 t^3}{3!}\right) + \log\left\{1 + r e^{ab + \frac{bt^2}{2!} + \frac{ct^3}{3!}}\right\} \end{aligned}$$

Further

$$(29)\quad \log\left\{1 + r e^{at + \frac{bt^2}{2!} + \frac{ct^3}{3!}}\right\} = \sum_{K=1}^{\infty}\frac{(-1)^{K-1}\cdot q^K}{K}\left(e^B - 1\right)^K + \log(1+r),$$

in which

$$B = at + \frac{bt^2}{2!} + \frac{ct^3}{3!} .$$

The right member of (29), may be put into the form

$$\left\{ \sum_{k=1}^{\infty} \sum_{i=1}^{k} \frac{(-1)^{1+i} \cdot q^k}{k} \binom{k}{i} \sum_{j=0}^{\infty} \frac{i^j B^j}{j!} \right\} + \log(1+r)$$

Therefore equating corresponding coefficients of $\frac{t^n}{n!}$ in the right and left members of (28), for $n > 3$, gives

(30)

$$L_n = \sum_{k=1}^{n} \sum_{i=1}^{k} \frac{(-1)^{1+i} \cdot q^k}{k} \binom{k}{i} \sum \frac{i^j \cdot n!}{\alpha! \beta! \gamma! \cdots (2!)^\beta (3!)^\gamma} \cdot (a)^\alpha (b)^\beta (c)^\gamma ,$$

where the last summation is taken over all values of $j$, such that the following diophantine equations are satisfied.

(31)
$$\alpha + \beta + \gamma = j,$$
$$\alpha + 2\beta + 3\gamma = n.$$

Using (30), I obtain, for $n = 1$ to 8 inclusive, the first eight semi-invariants of the compound as follows,

$$L_1 = m_1 + aq .$$

$$L_2 = \sigma_1^2 + q\left\{a^2(1-q) + b\right\}.$$

$$L_3 = {}_1\lambda_3 + a^3 q'_3 + 3abq'_2 + cq .$$

$$L_4 = a^4 q'_4 + 6a^2 b q'_3 + 3b^2 q'_2 + 4acq'_2 .$$

$$L_5 = a^5q'_5 + 10a^3bq'_4 + 15ab^2q'_3 + 10a^2cq'_3 + 10bcq'_2$$

$$(32)\ L_6 = a^6q'_6 + 15a^4bq'_5 + 45a^2b^2q'_4 + 15b^3q'_3 + 20a^3cq'_4 + 60abcq'_3 + 10c^2q'_2.$$

$$L_7 = a^7q'_7 + 21a^5bq'_6 + 105a^3b^2q'_5 + 105ab^3q'_4 + 35a^4cq'_5 + 210a^2bcq'_4 + 70ac^2q'_3 + 105b^2cq'_3$$

$$L_8 = a^8q'_8 + 28a^6bq'_7 + 210a^4b^2q'_6 + 420a^2b^3q'_5 + 105b^4q'_4 + 56a^5cq'_6 + 560a^3bcq'_5 + 280a^2c^2q'_4 + 840ab^2cq'_4 + 280bc^2q'_3.$$

in which $q'_2 \equiv q_2 = pq$, $q'_3 = pqq_3$, etc., and are in fact the same polynomials that occurred in the discussion of the case for normal components in *Section 1*.

The expression for the $L_n$ in (5) may be put into a form similar that that of (30), and then, if we compare these two forms, it is obvious that no new polynomials $q'_n$ will occur in addition to those which appeared in (5), and this would be true however many non-zero semi-invariants the two components may have.

Using the results established for the $L_n$ in Section 1, it is evident that, if we consider a particular semi-invariant, say $L_n$, of (5), the terms in the right member can be readily written down, if we determine all the $j$ part partitions of $n$, where $j$ and $n$ are fixed, using the integers 1 and 2 as part magnitudes. Suppose we have $\alpha$ parts, each equal to 1, and $\beta$ parts, each equal to 2, where $\alpha + \beta = j$ and $\alpha + 2\beta = n$, then such a partition corresponds to a term of the type $a^\alpha b^\beta q'_j$ (omitting the numerical coefficient), the factor $a^\alpha b^\beta$ arising from the last summation of (5), which

is clearly seen, if the latter be put into the same form as (30). In addition, it is to be noted that any $j$ part partition of $n$ will be unique for this case.

Now if we consider the case of the present *Section*, for $n>3$ the $L_n$ of (30) will be seen to contain, as well as the same terms of the corresponding $L_n$ of (5), some additional terms, the latter appearing on account of the fact that, since the integer 3 is now admitted along with 1 and 2 as a part magnitude, the $j$ part partitions of $n$ will no longer be unique for every possible value of $j$. These $j$ part partitions of $n$ will, moreover, give rise to terms of the type $a^{\alpha} b^{\beta} c^{\gamma} q'_j$ where the relations (31) are satisfied. Further, the total number of terms in a given $L_n$, for $n$ not too large, can be readily obtained by making use of the so-called "enumerating function," discussed in works on combinatorial analysis, which enables one to determine the number of partitions of a given integer $n$, when the number of parts $j$, and the part magnitudes 1, 2, 3, etc are fixed.

It would appear, from the above discussion that the partition method of obtaining the terms of $L_n$ could be carried over to the most general case, in which the components may possess non-zero semi-invariants of all orders.

I shall now indicate, without going into detail, that, if in the expressions (30) for $n>3$, I set every $q'_j$ equal to unity, then the $L_n$ become functions of a, b, and c only, which I call $\beta_n$, and the latter obey a recursion law analogous to the one established for the $\beta_s$ of (10), namely

$$\beta_{n+1} = \left(a + b\frac{\partial}{\partial a} + c\frac{\partial}{\partial b}\right)\beta_n, \tag{33}$$

where now

$$\beta_n = \frac{d^n}{dt^n}\left\{e^{at+\frac{bt^2}{2!}+\frac{ct^3}{3!}}\right\}\Bigg|_{t=0} \quad \text{by definition.}$$

Putting $\phi(t) = at + \frac{bt^2}{2!} + \frac{ct^3}{3!}$

(34) $\frac{d^n}{dt^n}\left\{e^{\phi(t)}\right\} = P_n\left(c, b+ct, a+bt+\frac{ct^2}{2}\right) e^{\phi(t)}$,

where $P_n(x, y, z)$ is an $n$-th degree polynomial in x, y, and z. Again, it is readily seen that

(35)
$$e^{\phi(t)} \cdot \frac{d}{dt}\left\{P_n\left(c, b+ct, a+bt+\frac{ct^2}{2}\right)\right\}\Bigg|_{t=0}$$
$$= \left(b \cdot \frac{\partial}{\partial a} + c\frac{\partial}{\partial b}\right)\left\{P_n\left(c, b+ct, a+bt+\frac{ct^2}{2}\right) \cdot e^{\phi(t)}\Bigg|_{t=0}\right\},$$

and that

(36) $$P_n \cdot \frac{d}{dt}\left\{e^{\phi(t)}\right\}\Bigg|_{t=0} = a \cdot P_n \cdot e^{\phi(t)}\Bigg|_{t=0}.$$

Now deriving the left member of (34), with respect to $t$, and then setting $t=0$ gives the next $\beta$, viz. $\beta_{n+1}$, whilst the derivative of the right member of (34), then setting $t=0$, would equal the sum of the right members of (35) and (36), and thus the law in this case is established.

It is at once apparent that the recursion formulae of (13) and (33) may be generalized, so that, if the two components should possess non-zero semi-invariants of all orders, the law for the $\beta_n$ would then be

$$\beta_{n+1} = \left(a + b\frac{\partial}{\partial a} + c\frac{\partial}{\partial b} + d\frac{\partial}{\partial c} + \text{ etc., } \ldots\right)\beta_n.$$

a, b, c, d, etc. being the differences between the 1st, 2nd, etc. semi-invariants of the two components, respectively. Thus it appears that the actual writing down of the expressions for the parameters of a compound frequency function in terms of the parameters of its two components may be reduced to a partititon process, and a taking of derivatives.

*Section 5. The limiting compound frequency function, when the number of components is allowed to become indefinitely large.*

It is to be noted that if the compound is assumed to be composed of a greater number of components than two, then the mathematical development beoomes heavy, but a rather interesting case arises, when we consider the form of the limiting compound frequency function, when its components, infinite in number, and identical in form, each contribute the same proportion to the total frequency of the compound, and have their means distributed according to the known frequency law $f(x)$.

First of all I consider the compound to be composed of a finite number of components, say $M+1$, of the type indicated, and later pass to the limit, allowing $M$ to become indefinitely large.

I write now

$$(37)\quad e^{L_1 t + L_2 \frac{t^2}{2!} + L_3 \frac{t^3}{3!} + \cdots} = \sum_{i=1}^{M+1} p_i \cdot e^{\lambda_{i1} t + \lambda_{i2} \frac{t^2}{2!} + \lambda_{i3} \frac{t^3}{3!} + \cdots}$$

in which $p_i = \frac{1}{M+1}$ for all $i = 1, 2, 3, \ldots$ etc. and $\lambda_{i1}, \lambda_{i2}$, etc, are the 1st, 2nd, etc. semi-invariants respectively of the $i$-th component. The right member of (37) may be written

$$e^{\lambda_{11} t + \lambda_{12} \frac{t^2}{2!} + \lambda_{13} \frac{t^3}{3!} + \cdots} \cdot \frac{1}{M+1} \left\{ 1 + \sum_{i=1}^{M} e^{(m_{i+1} - m_1) t} \right\},$$

in which $m_1$ is the mean of some component and $m_{i+1}$ is the mean of the $(i+1)$st component.

If we now assume that $m_1 = \lambda_{11} = 0$, then

(38)
$$e^{L_1 t + L_2 \frac{t^2}{2!} + L_3 \frac{t^3}{3!} + \cdots} = e^{\lambda_{12} \frac{t^2}{2} + \lambda_{13} \frac{t^3}{3!} + \cdots} \cdot \frac{1}{M+1} \left\{ 1 + \sum_{i=2}^{M+1} e^{m_i t} \right\},$$

and the right member of the last relation is the generating function for the moments of the compound frequency function. Allowing $M$ to become indefinitely large, we have

$$\lim_{M \to \infty} \frac{1}{M+1} \left\{ 1 + e^{m_2 t} + e^{m_3 t} + \cdots \cdots e^{m_{M+1} t} + \cdots \right\}$$
$$= \int_{-\infty}^{\infty} e^{xt} \cdot f(x)\, dx = G_x(t).$$

Therefore the limit of the generating function (the right member of (38) is given by

$$e^{\lambda_{12} \frac{t^2}{2} + \lambda_{13} \frac{t^3}{3!} + \cdots} \cdot G_x(t),$$

so that the semi-invariants $L_n$ of the limiting compound frequency function are given by

$$e^{L_1 t + L_2 \frac{t^2}{2!} + L_3 \frac{t^3}{3!} + \cdots} = e^{\lambda_{12} \frac{t^2}{2!} + \lambda_{13} \frac{t^3}{3!} + \cdots} \cdot G_x(t).$$

From this last relation, it will be seen that the mean of the limiting compound is equal to the mean of the means of the components. Further, if the means of the components are normally distributed, and the components themselves are normal, then the limiting compound frequency function is also normal. More generally, if the components are normal, and their means follow any frequency law $f(x)$, then the limiting compound function also follows this same law. If now, considering the most general case of all, where the components may have non-zero semi-invariants of all orders, and the means of the components are distributed according to the frequency law $F(x)$, then the semi-invariants of the limiting compound frequency function may always be calculated, and will be given by

$$L_k = \lambda_{1k} + \ell_k,$$

in which $L_k$, $\lambda_{1k}$, $\ell_k$, are the $k$-th semi-invariants of the limiting compound function, of one of the components, and of $F(x)$ respectively.

This shows that the variate $z$ of the limiting compound frequency function, is distributed as if it were the sum of two independent variates, one of which is distributed according to the law of the means, and the other according to one of the components. To write down the actual distribution function for the limiting compound is quite another matter, but since we may write, in the limit, when $M \to \infty$,

$$e^{L_1 t + L_2 \frac{t^2}{2!} + L_3 \frac{t^3}{3!} + \cdots} = e^{\lambda_{12} \frac{t^2}{2!} + \lambda_{13} \frac{t^3}{3!} + \cdots} \cdot e^{\ell_1 t + \ell_2 \frac{t^2}{2!} + \cdots\cdots}$$

then, the distribution function sought, provided it fulfills the necessary conditions, may be given formally by means of the Fourier Integral Theorem.

## PART 2.

As indicated previously, in the introduction to this paper, we are concerned in this second part with an entirely new problem, in which we are now sampling from two distinct parent populations instead of from only one, as in Part 1. Hence, in order to obtain the desired sampling results, we must have recourse to an entirely different method of treatment from any we have made use of heretofore. I shall suppose that $\phi_1(x), \phi_2(x)$, the two parent populations may possess non-zero semi-invariants of all orders, and that a random sample of $r$ is taken from the first population, and a random sample of $s$ from the second population, these two samples being then combined to give the composite sample from the combined populations.

*Section 6. The semi-invariants of "moments about a fixed point," in samples from the compound frequency function.*
With the above hypotheses, I shall derive in this section, expressions for the semi-invariants of "moments about a fixed point" in samples from the compound population.

Calling the required semi-invariants $S_1$, $S_2$, $S_3$. arc, we have, by definition

$$e^{S_1 t + S_2 \frac{t^2}{2!} + S_3 \frac{t^3}{3!} + \cdots}$$

(39)

$$= \int_{-\infty}^{\infty} \int [\phi_1(x)]^r [\phi_2(x)]^s \cdot e^{\left\{ \sum_{i=1}^{r} {}_1x_i^n + \sum_{j=r+1}^{r+s} {}_2x_j^n \right\} \frac{t}{N}} (dx)^{r+s}$$

in which $\phi_1(x)$, $\phi_2(x)$ are the initial parent frequency functions, and ${}_1x_i$, ${}_2x_j$, indicates that the variate was taken from the first and second parent respectively. By a suitable transformation of the

parameter in the power of the exponential which appears in the right member of (39), this same member may be put into the required form

$$\int_{-\infty}^{\infty}\!\!\int [\varphi_1(x)]^r \cdot e^{\frac{rt}{N}\left\{\frac{1}{r}\sum_{i=1}^{r} {}_1x_i^n\right\}} \cdot (dx)^r \cdot \int_{-\infty}^{\infty}\!\!\int [\varphi_2(x)]^s \cdot e^{\frac{st}{N}\left\{\frac{1}{s}\sum_{j=r+1}^{r+s} {}_2x_j^n\right\}} (dx)^s.$$

On equating corresponding coefficients of $\frac{t^k}{k!}$ in the last expression and the left member of (39), I get

$$S_k(V_n') = \frac{r^k \cdot S_k'(V_n') + s^k \cdot S_k''(V_n')}{(r+s)^k},$$

in which $S_1'(V_n')$, $S_2'(V_n')$, $S_3'(V_n')$, etc., and $S_1''(V_n')$, $S_2''(V_n')$, $S_3''(V_n')$, etc., are the 1st, 2nd, 3rd, etc., semi-invariants for $V_n'$, in samples from the two component populations $\varphi_1(x)$, and $\varphi_2(x)$, respectively, the values of which are well known.[1]

*Section 7. The semi-invariants of "moments about the mean" in samples from the compound frequency function.*

Employing the same sampling procedure as in the last section, I wish now to consider the semi-invariants of "moments in samples from the combined population, about the mean of the combined sample" and to express them in terms of $\alpha_1$, $\alpha_2$, and $\beta_1$, $\beta_2$, (the semi-invariants if the component distributions $\varphi_1(x)$, $\varphi_2(x)$ respectively), and $r$ and $s$.

In order to obtain the desired results, I have made use of a

[1] Loc. Cit. pp. 12-13.

modification and extension of a method originally employed by C. C. Craig[1] for the case of sampling from one normal parent population. I shall first develop the theory for my case, on the basis that the two parent populations may possess non-zero semi-invariants of all orders, imposing the condition of normality only when actually computing the desired results. The mean of the combined sample is

$$V_1' = \sum_{i=1}^{r} {}_1x_i + \sum_{j=r+1}^{r+s} {}_2x_j$$

We wish to find the semi-invariants $S_k$ of $V_n \equiv \sum_{i=1}^{N} \frac{\delta_i^n}{N}$ in which $\delta_i = x_i - V_1'$ and $r+s=N$, for particular values of $n$, and for infinitely many sets of $r+s$ variates, assuming that each member of each set is independent of all the rest. The $N$ $\delta$'s in each set satisfy $\sum_{i=1}^{N} \delta_i = 0$ .

Now, let $F(\delta_1, \delta_2 \cdots \delta_{N-1})$ be the correlation function of the first $N-1$ $\delta$'s. Then $F(\delta_1, \delta_2 \cdot \delta_{N-1}) d\delta_1 d\delta_2 \cdot d\delta_{N-1}$ gives the probability that the first $N-1$ $\delta$'s fall simultaneously within a cell

$$\left(\delta_1 \pm \tfrac{1}{2} d\delta_1, \delta_2 \pm \tfrac{1}{2} d\delta_2, \cdot \quad \cdots, \delta_{N-1} \pm \tfrac{1}{2} d\delta_{N-1}\right)$$

The semi-invariants of $F(\delta_1, \delta_2 \cdots \delta_{N-1})$ are defined by

$$(40) \quad e^{\left(\sum_{i=1}^{N-1} \lambda_i t_i\right) + \frac{1}{2}\left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(2)} + \frac{1}{3!}\left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(3)} + \cdots}$$

$$= \int d\delta_1 \int d\delta_2 \int_{-\infty}^{\infty} \cdots \int d\delta_{N-1} \cdot F(\delta_1, \delta_2 \cdot \delta_{N-1}) \cdot e^{\sum_{i=1}^{N-1} \delta_i t_i}$$

$$= 1 + \left(\sum_{i=1}^{N-1} V_i t_i\right) + \frac{1}{2!}\left(\sum_{i=1}^{N-1} V_i t_i\right)^{(2)} + \frac{1}{3!}\left(\sum_{i=1}^{N-1} V_i t_i\right)^{(3)} + \cdots$$

[1] Loc. Cit. pp. 1 to 35.

where e.g.

$$(41) \quad \left(\sum_{i=1}^{2} \lambda_i t_i\right)^{(2)} = \lambda_{20} t_1^2 + 2\lambda_{11} t_1 t_2 + \lambda_{02} t_2^2 ;$$

Setting

$$\delta_i = \sum_{j=1}^{N} a_{ij} x_j \qquad a_{ij} = -\frac{1}{N},\ i \neq j$$

$$a_{ii} = \frac{N-1}{N}$$

we have

$$e^{\left(\sum_{i=1}^{N-1} \lambda_i t_i\right) + \frac{1}{2!}\left(\sum_{i=1}^{N-1} \lambda_i t_i\right)^{(2)} + \cdots\cdots}$$

$$= \int dx_1 \int_{-\infty}^{\infty} dx_2 \int \cdots \int dx_N \left[\phi_1(x)\right]^r \left[\phi_2(x)\right]^s \cdot e^{\sum_{i=1}^{N-1} \sum_{j=1}^{N} a_{ij} x_j t_i}$$

$$= e^{\alpha_1\left(\sum_{i=1}^{N-1} a_{i1} t_i\right) + \frac{\alpha_2}{2!}\left(\sum_{i=1}^{N-1} a_{i1} t_i\right)^2 + \cdots \quad \cdots}$$

$$e^{\alpha_1\left(\sum_{i=1}^{N-1} a_{i2} t_i\right) + \frac{\alpha_2}{2!}\left(\sum_{i=1}^{N-1} a_{i2} t_i\right)^2 + \cdots\cdots}$$

$$\vdots$$

$$e^{\alpha_1\left(\sum_{i=1}^{N-1} a_{ir} t_i\right) + \frac{\alpha_2}{2!}\left(\sum_{i=1}^{N-1} a_{ir} t_i\right)^2 + \cdots \quad \cdots}$$

$$e^{\beta_1\left(\sum_{i=1}^{N-1} a_{i,r+1} t_i\right) + \frac{\beta_2}{2!}\left(\sum_{i=1}^{N-1} a_{i\,r+1} t_i\right)^2 + \cdots}$$

$$e^{\beta_1\left(\sum_{i=1}^{N-1} a_{i,r+2} t_i\right) + \frac{\beta_2}{2!}\left(\sum_{i=1}^{N-1} a_{i,r+2} t_i\right)^2 + \cdots\cdots}$$

$$\vdots$$

$$e^{\beta_1\left(\sum_{i=1}^{N-1} a_{i,r+s} t_i\right) + \frac{\beta_2}{2!}\left(\sum_{i=1}^{N-1} a_{i,r+s} t_i\right)^2 + \cdots\cdots}$$

in which $\alpha_1, \alpha_2$, etc., are the 1st, 2nd, etc. semi-invariants of the first component $\phi_1(x)$, and $\beta_1, \beta_2$, etc. are the corresponding semi-invariants for the second component $\phi_2(x)$. It follows then, that

$$\left(\sum_{i=1}^{N-1}\lambda_i t_i\right)^{(k)} = \alpha_k\left\{\left(\sum_{i=1}^{N-1}a_{i1}t_i\right)^k + \left(\sum_{i=1}^{N-1}a_{i2}t_i\right)^k + \cdot \quad + \left(\sum_{i=1}^{N-1}a_{ir}t_i\right)^k\right\}$$

$$+\beta_k\left\{\left(\sum_{i=1}^{N-1}a_{i,r+1}t_i\right)^k + \left(\sum_{i=1}^{N-1}a_{i,r+2}t_i\right)^k + \cdot + \left(\sum_{i=1}^{N-1}a_{i,r+s}t_i\right)^k\right\}$$

or

$$\lambda_{k_1k_2\cdots k_{N-1}} = \alpha_k\left\{a_{11}^{k_1}a_{21}^{k_2}\cdot \quad \cdot a_{N-1,1}^{k_{N-1}} + a_{12}^{k_1}a_{22}^{k_2}\cdots a_{N-1,2}^{k_{N-1}} + \cdot + a_{1r}^{k_1}a_{2r}^{k_2}\cdot \; a_{N-1,r}^{k_{N-1}}\right\}$$

$$+\beta_k\left\{a_{1,r+1}^{k_1}a_{2,r+1}^{k_2}\cdots a_{N-1,r+1}^{k_{N-1}} + a_{1,r+2}^{k_1}a_{2,r+2}^{k_2}\cdot a_{N-1,r+2}^{k_{N-1}} + \cdot\right.$$

$$\left. + a_{1,r+s}^{k_1}a_{2,r+s}^{k_2}\cdot \; a_{N-1,r+s}^{k_{N-1}}\right\},$$

where $k_1 + k_2 + \quad \cdots + k_{N-1} = k$.

so that

$$\lambda_{k_1k_2\cdots k_{N-1}} = \alpha_k\left\{\sum_{j=1}^{r}a_{1j}^{k_1}a_{2j}^{k_2}\cdot \quad \cdot a_{N-1,j}^{k_{N-1}}\right\}$$

$$+\beta_k\left\{\sum_{j=r+1}^{r+s}a_{1j}^{k_1}a_{2j}^{k_2}\cdots a_{N-1,j}^{k_{N-1}}\right\},$$

By substituting for the a's from the relation (41) in the last equation, the latter may be reduced to the following convenient form, where, now, $\lambda_{k_1k_2\cdots k_{N-1}} = \lambda_{k_1k_2\cdots\cdots k_N}$, provided that, at least one of the $k_i$ in $\lambda_{k_1k_2\cdots k_N}$ is equal to zero.

$$\lambda_{k_1k_2\cdots k_N} = \frac{1}{N^k}\left\{\alpha_k\left(\sum_{i=1}^{r}(-1)^{k-k_i}(N-1)^{k_i}\right)\right. \tag{42}$$

$$\left. + \beta_k\left(\sum_{i=r+1}^{r+s}(-1)^{k-k_i}(N-1)^{k_i}\right)\right\}.$$

In the case of one parent population, all $\lambda_{k_1k_2\cdots k_N}$ of the same type, i.e. whose subscripts $k_1k_2\cdots k_N$ are merely different permutations of the same set of integers $k_1, k_2, \ldots, k_N$, were equal. This is no longer true for two parent populations, for we must now distinguish between the $\delta$'s which arise from observations from the population $\phi_1(x)$, and those from $\phi_2(x)$. I therefore introduce, at this point, what I shall call the "bar notation". For example, from the relation (42), all $\lambda_{k_1k_2\cdots k_r|k_{r+1}\cdots k_{r+s}}$ will be of the same type, and therefore equal, if the first $r$ subscripts are merely different permutations of the same set of integers $k_1, k_2, \cdots k_r$ whilst, quite apart from the first $r$ subscripts, the last $s$ subscripts are also different permutations of the same set of integers $k_{r+1}$, $k_{r+2}, \ldots k_{r+s}$. In writing down the semi-invariants of the correlation function $F$, using the "bar notation", for convenience, I shall suppress zero subscripts. Further, on account of the relation (42), all $\lambda_{k_1k_2\cdots k_r|k_{r+1}\cdots k_{r+s}}$ for which $k_1+k_2+\cdots k_r+k_{r+1}+\cdots+k_{r+s} \geq 3$ will vanish, since we are assuming now, that our two parent populations are normal. As a matter of fact, the only $\lambda_{k_1k_2\cdots k_r|k_{r+1}\cdots k_{r+s}}$ that I shall require here are the following,

$$\lambda_{1|0} = \frac{1}{N}\left\{N\alpha_1\right\}.$$

$$\lambda_{0|1} = \frac{1}{N}\left\{N\beta_1\right\}.$$

$$\lambda_{2|0} = \frac{1}{N^2}\left\{r\alpha_2 + s\beta_2 + N(N-2)\alpha_2\right\}.$$

$$\lambda_{0|2} = \frac{1}{N^2}\left\{r\alpha_2 + s\beta_2 + N(N-2)\beta_2\right\}.$$

$$\lambda_{11|0} = \frac{1}{N^2}\left\{r\alpha_2 + s\beta_2 - 2N\alpha_2\right\}.$$

$$\lambda_{0|11} = \frac{1}{N^2}\left\{r\alpha_2 + s\beta_2 - 2N\beta_2\right\}.$$

$$\lambda_{1|1} = \frac{1}{N^2}\left\{r\alpha_2 + s\beta_2 - N(\alpha_2+\beta_2)\right\}.$$

The above expressions were obtained from (42), after assuming, (without loss of generality,) that $r\alpha_1 + s\beta_1 = 0$. If, instead of this last assumption, I had assumed that $\lambda_{1|0}$ or $\lambda_{0|1}$ were equal to zero, not only would the symmetry of the final results have been destroyed, but the amount of labour necessary to obtain them would also have been doubled. The symmetric substitution actually made, required that only half the final number of terms be obtained, the remaining half in any particular result being readily written down by interchanging the $\alpha$'s and $\beta$'s as well as $r$ and $s$.

Now, let $P(V_n)$ be the probability function for $V_n \equiv \sum_{i=1}^{N} \frac{\delta_i^n}{N}$

The semi-invariants of $P(V_n)$ are then defined by

$$
(43)\qquad
\begin{aligned}
e^{S_1 t + S_2\frac{t^2}{2!} + S_3\frac{t^3}{3!}+\cdots} &= \int_{-\infty}^{\infty} P(V_n)\, e^{V_n t}\cdot dV_n \\
&= \int d\delta_1 \int_{-\infty}^{+\infty} d\delta_2 \cdots \int d\delta_N \cdot F(\delta_1,\delta_2 \cdots \delta_N)\, e^{\sum_{i=1}^{N}\delta_i^n \frac{t}{N}}
\end{aligned}
$$

Regarding the use of $F(\delta_1,\delta_2,\cdots\delta_N)$ instead of $F(\delta_1,\delta_2,\cdots\delta_{N-1})$ in the above relation, see paper by C. C. Craig.[1]

We wish now to express the semi-invariants $S_K$ in terms of the semi-invariants $\lambda_{K_1K_2\cdots K_N}$ of the correlation function $F(\delta_1,\delta_2\cdots\delta_N)$ for the $\delta$'s. The semi-invariants $L_{rst}\ldots$ of the correlation function for $\delta_i^n$ are defined by

$$
(44)\qquad
\begin{aligned}
e^{\left(\sum_{i=1}^{N}L_i t_i\right)+\frac{1}{2!}\left(\sum_{i=1}^{N}L_i t_i\right)^{(2)}+\frac{1}{3!}\left(\sum_{i=1}^{N}L_i t_i\right)^{(3)}+\cdots} &= \int d\delta_1\int d\delta_2\int_{-\infty}^{+\infty}\cdots\int d\delta_N\, F(\delta_1,\delta_2\cdots\delta_N)\, e^{\sum_{i=1}^{N}\delta_i^n t_i} \\
&= 1+\left(\sum_{i=1}^{N}V_{ni}t_i\right)+\left(\sum_{i=1}^{N}V_{ni}t_i\right)^{(2)}+\cdots
\end{aligned}
$$

by expansion of the exponential function. Then, comparing the

[1] Loc. Cit. pp. 18 to 19

relations in (43) and (44). it is readily seen that

$$(45) \qquad S_k = \frac{1}{N^k} \sum \frac{k!}{k_1!\, k_2! \,\cdots\, k_r!} \cdot L_{k_1 k_2 \cdots k_r}$$

in which the summation is taken over all values of $k_1, k_2, \ldots k_r$, such that

$$k_1 + k_2 + \cdots + k_r = k .$$

Making use of the explicit relations for semi-invariants in terms of moments and vice versa, we have from (44) and (40)

$$(46) \qquad \left(\sum_{i=1}^{N} L_i t_i\right)^{(k)} = \sum\sum \cdots \sum \frac{(-1)^{(r+s+t+\cdots)-1}\left[(r+s+t+\cdots)-1\right]!\,k!}{(a!)^r (b!)^s (c!)^t \cdots} \times \frac{\left[\left(\sum_{i=1}^{N} \nu_{ni} t_i\right)^{(a)}\right]^r \left[\left(\sum_{i=1}^{N} \nu_{ni} t_i\right)^{(b)}\right]^s \cdots}{r!\, s!\, t!}$$

$$(47) \qquad \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(n)} = \sum\sum \cdots \sum \frac{n!\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(a)}\right]^r \cdot \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(b)}\right]^s \cdots}{(a!)^r (b!)^s (c!)^t \cdots \; r!\, s!\, t!}$$

where, in both these relations

$$a > b > c > \cdots$$

and

$$ar + bs + ct + \cdots = n$$

From the relations (46) and (47), the $L_n$ can be found in terms of the moments of $F$, and these, in turn, can be found in terms

of the semi-invariants of $F$, by equating the coefficients of like powers of the $t$'s on both sides of the two equations. Examples of the kind of relations obtained from (46) and (47) in particular cases would be as follows,

$$\left(\sum_{i=1}^{N} L_i t_i\right)^{(3)} = \left(\sum_{i=1}^{N} V_{2i} t_i\right)^{(3)} - 3\left(\sum_{i=1}^{N} V_{2i} t_i\right)^{(2)}\left(\sum_{i=1}^{N} V_{2i} t_i\right) + 2\left(\sum_{i=1}^{N} V_{2i} t_i\right).$$

Therefore

$$L_{210\ldots 0} = V_{420.\,0} - V_{40\ldots 0} \cdot V_{020.\,0}$$

$$- 2V_{220..\,0}\, V_{20..\,0} + 2V_{20\ 0}^{2}\, V_{020\,.0},$$

$$\left(\sum_{i=1}^{N} v_i t_i\right)^{(6)} = \left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(6)} + 15\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(4)}\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}$$

$$+ 10\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(3)}\right]^2 + 15\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^3.$$

Therefore

$$V_{420\ 0} = \lambda_{420\,.\,0} + \lambda_{40\ 0} \cdot \lambda_{020\ 0} + 8\lambda_{310\,.0}\lambda_{110.\,.0}$$

$$+ 6\lambda_{220..\,0}\,\lambda_{20..0} + 6\lambda_{210\,.\,0}^{2} + 4\lambda_{30\,..0}\,\lambda_{120.\,.0}$$

$$+ 3\lambda_{20\,.\,0}^{2} \cdot \lambda_{020\ldots 0} + 12\lambda_{20..\,0} \cdot \lambda_{110\ldots 0}^{2}.$$

In my work, I actually make use of the following relations obtained from (47), with certain terms omitted, which vanish when each of the parent distributions is normal.

$$
\begin{aligned}
&\text{(i)} && \left(\sum_{i=1}^{N} \nu_i t_i\right) = \left(\sum_{i=1}^{N} \lambda_i t_i\right). \\
&\text{(ii)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(2)} = \left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)} + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(1)}\right]^2. \\
&\text{(iii)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(3)} = 3\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\left(\sum_{i=1}^{N} \lambda_i t_i\right) + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^3. \\
&\text{(iv)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(4)} = 3\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^2 + 6\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^2 + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^4. \\
&\text{(v)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(5)} = 15\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^2\left(\sum_{i=1}^{N} \lambda_i t_i\right) \\
& && \qquad + 10\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^3\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)} + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^5 \\
(48) \quad &\text{(vi)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(6)} = 45\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^2\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^2 \\
& && \qquad + 15\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^4\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)} + 15\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^3 + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^6. \\
&\text{(vii)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(7)} = 105\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^2\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^3 \\
& && + 21\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^5 + 105\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^3\left(\sum_{i=1}^{N} \lambda_i t_i\right) + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^7. \\
&\text{(viii)} && \left(\sum_{i=1}^{N} \nu_i t_i\right)^{(8)} = 210\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^2\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^4 + 28\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)} \cdot \times \\
& && \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^6 + 420\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^3\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^2 + 105\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}\right]^4 + \left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^8.
\end{aligned}
$$

By substituting the expressions for the $L$'s given in (46), into the right member of (45), a direct expression for the $S_K$'s in terms of the moments of $F$ is obtained, viz.

$$S_k = \frac{1}{N^k}\Sigma\ldots\Sigma L_{k_1 k_2 \ldots k_r} \frac{k!}{k_1!\,k_2!\ldots k_r!} = \frac{1}{N^k}\left(\sum_{i=1}^{N} L_i\right)^{(k)}.$$

omitting the parameters $t_i$, and in which $\left(\sum_{i=1}^{N} L_i\right)^{(k)}$ is given by (46). From relation (43) to this point, the theory follows exactly that given in the paper already cited.

I shall next quote the final expressions A for some of the $S_k(v_n)$ obtained by C. C. Craig,[1] in terms of the moments of $F$, for the case of one parent population, and then I shall write down the modified expressions B, when two parent populations are involved.

A. 1. $S_1(V_n) = \frac{1}{N}\left\{N V_{n,0}\right\}.$

2. $S_2(V_n) = \frac{1}{N^2}\left\{N V_{2n,0} + N(N-1)V_{n,n,0} - N^2 V^2_{n,0}\right\}.$

B. 1. $S_1(V_n) = \frac{1}{N}\left\{r V_{n|0} + s V_{0|n}\right\}$

$$2.\ S_2(V_n) = \frac{1}{N^2}\Big\{r V_{2n|0} + s V_{0|2n} + V(r-1)V_{n,n|0} + s(s-1)V_{0|n,n}$$
$$+ 2rs V_{n|n}\Big\} - \left[r^2 V^2_{n|0} + s^2 V^2_{0|n} + 2rs V_{n|0}\cdot V_{0|n}\right].$$

In the paper mentioned above, expressions were also derived, using a method similar to the one already indicated, for the semi-invariants of the correlation function of two moments about the mean. I have made use of a modification of only one of these expressions as follows,

[1] Loc. Cit. p. 22.

$$\text{A.} \quad S_{11}(V_m, V_n) = \frac{1}{N^2}\left\{N V_{m+n,0} + N(N-1)V_{m,n,0} - N^2 V_{m,0} V_{n,0}\right\}$$

$$\text{B.} \quad S_{11}(V_m, V_n) = \frac{1}{N^2}\Big\{r V_{m+n|0} + s V_{0|m+n} + r(r-1)V_{m,n|0}$$

$$+ s(s-1)V_{0|m,n} + rs V_{m|n} + rs V_{n|m}$$

$$-\left[r^2 V_{m|0} V_{n|0} + s^2 V_{0|m} V_{0|n} + rs V_{m|0} V_{0|n} + rs V_{0|m} V_{n|0}\right]\Big\}.$$

Here again, for the moments of the correlation function, I employ the "bar notation", its meaning being exactly the same as in the case of the $\lambda_{k_1 k_2 \cdots k_r | k_{r+1} \cdots k_{r+s}}$, the discussion regarding identical types of moments and their equality, corresponding also in every detail. Once more, zero subscripts are suppressed.

It now becomes necessary to express the modified moments in the expressions B, for particular values of $m$ and $n$, in terms of the $\lambda_{k_1 k_2 \cdots k_r | k_{r+1} \cdots k_{r+s}}$ of the correlation function $F$. To this end, I make use of the relations (48), in conjunction with the so-called "$D_s$ operator of Hammond"[1] which splits off a total integral part $s$, made up by addition from any or all of a permutation of integers.

At this point also, it is necessary to modify somewhat the use of the $D_s$ operator, because of the "bar notation" used to designate the moments and the semi-invariants of the correlation function $F$, when two parent populations are being considered. In making up the total integral part $s$, split off from the permutation of integers

[1] MacMahon—Combinatorial Analysis, Vol. I, p 27

$k_1 k_2 \ldots k_r | k_{r+1} \cdot \cdot k_{r+s}$, the parts split off from the set of integers $k_1 k_2 \cdot \cdot k_r$ must be kept distinct from those parts which are split off from the set of integers $k_{r+1}, k_{r+2}, \ldots k_{r+s}$, and this same rule applies also to the residual permutations from each of these two sets, after all the parts, with sum $s$ have been finally split off. Hence, the use of the "bar notation" to effect this distinction. To illustrate exactly what is meant here, suppose that I wish to express $\nu_{3|2}$ in terms of $\lambda_{k_1 k_2 \cdots k_r | k_{r+1} \cdots k_{r+s}}$. In this case, I shall use the relation (v) of the set of equations (48), and I shall merely consider the contribution made by the second term in the right member of (v) to the final expression for $\nu_{3|2}$, the other terms of (v) being treated in a similar manner. Now $\nu_{3|2}$ (omitting a numerical factor) is the coefficient of $t_1^3 t_2^2$ in the left member of (v). I therefore seek the corresponding coefficient of $t_1^3 t_2^2$ in the second term of the right member of (v) this term being $10\left[\left(\sum_{i=1}^{N} \lambda_i t_i\right)\right]^3 \left(\sum_{i=1}^{N} \lambda_i t_i\right)^{(2)}$. Using the modified form of the $D_s$ operator, we have

$$D_2 D_1^3 (3|2) = (2|0) D_1^2 (1|2) + (0|2) D_1^2 (3|0) + (1|1) D_1^2 (2|1)$$

$$= 3(2|0)(1|0)(0|1)^2 + (0|2)(1|0)^3 + 3(1|1)(1|0)^2(0|1)$$

Now, we are able to write down immediately the terms in $\lambda_{k_1 k_2 \cdots, k_r | k_{r+1} \cdots \cdot k_{r+s}}$ which arise. They are

$$(49) \qquad 3\lambda_{2|0}\lambda_{1|0}\lambda_{0|1}^2 + \lambda_{0|2}\lambda_{1|0}^3 + 6\lambda_{1|1}\lambda_{1|0}^2\lambda_{0|1}$$

Ordinarily, the numerical coefficients in (49), will need to be multiplied by an integral factor, obtained as follows. A term $t_1^3\, t_2^2$ may be chosen from the expansion of $\left(\sum_{i=1}^{N} v_i t_i\right)^{(5)}$ in $\frac{5!}{3!\,2!}$ or 10 (in general, in $C_1$) ways. The numerical coefficient of the second term in the right member of (V) is also 10 (and in general, is, say $C_2$). The required factor for the above example is unity (or, in general, the quotient $\frac{C_2}{C_1}$. It should be noticed in addition, that the sum of the coefficients in the final expression (49) should equal the numerical coefficient of the second term in the right member of (V), with which we started, and this is seen to be the case.

As a check, one may observe, that if in the results which I obtain for $S_k(v_n)$, $S_{k\ell}(v_m, v_n)$, the two normal parent populations are identified, then the results for a single normal parent population are obtained. Note that, to get the results as usually given for a single normal parent population, one would further have to set $\alpha_1 = \beta_1 = 0$.

I derived the following results, which have been checked by calculating the corresponding results for a single normal parent population, without assuming that the first order semi-invariants of the type $\lambda_{100\cdots00}$ are equal to zero.

$$s_1(v_2) = \frac{1}{N^2}\left\{(N-1)(r\alpha_2 + s\beta_2) + N(r\alpha_1^2 + s\beta_1^2)\right\}.$$

$$s_1(v_3) = \frac{1}{N^2}\left\{3(N-2)(r\alpha_1\alpha_2 + s\beta_1\beta_2) + N(r\alpha_1^3 + s\beta_1^3)\right\}.$$

$$s_1(v_4) = \frac{1}{N^3}\left\{\left[3Nr\left\{N(N-2)^2 + r(2N-3)\right\}\right]\alpha_2^2 + \left[3Ns\left\{N(N-2)^2 + s(2N-3)\right\}\right]\beta_2^2\right.$$

$$+N^4(r\alpha_1^4+s\beta_1^4)+6N^2\left[r(N^2-2N+r)\alpha_1^2\alpha_2+s(N^2-2N+s)\beta_1^2\beta_2\right]$$

$$+\left[6Nrs(2N-3)\right]\alpha_2\beta_2+6N^2rs\left[\alpha_1^2\beta_2+\alpha_2\beta_1^2\right]\Big\}$$

$$s_2(v_2)=\frac{2}{N^4}\Big\{r(N^2-2N+r)\alpha_2^2+s(N^2-2N+s)\beta_2^2+2rs\Big[(r\alpha_1^2\beta_2+s\alpha_2\beta_1^2)$$

$$+\alpha_2\beta_2+(s\alpha_1^2\alpha_2+r\beta_1^2\beta_2)-2(s\alpha_1\alpha_2\beta_1+r\alpha_1\beta_1\beta_2)\Big]\Big\}.$$

$$s_2(v_3)=\frac{3}{N^6}\Big\{6r\left[N^4-N^3+N^2s^2-s(r^2+5Ns)\right]\alpha_1^2\alpha_2^2+6s\Big[N^4-N^3+N^2r^2$$

$$-r(s^2+5Nr)\Big]\beta_1^2\beta_2^2+3N^2rs\Big[(s\alpha_1^4\alpha_2+r\beta_1^4\beta_2)+(r\alpha_1^4\beta_2+s\alpha_2\beta_1^4)$$

$$-2(s\alpha_1^2\alpha_2\beta_1^2+r\alpha_1^2\beta_1^2\beta_2)\Big]+3rs\Big[4(s^2\alpha_1\alpha_2^2\beta_1+r^2\alpha_1\beta_1\beta_2^2)$$

$$+8rs(\alpha_1\alpha_2\beta_1\beta_2)-2\left\{s(N^2-2N-r)\alpha_2^2\beta_1^2+r(N^2-2N-s)\alpha_1^2\beta_2^2\right\}$$

$$+2\left\{\left[N(6-N)(r-s)+2r^2\right]\alpha_2\beta_1^2\beta_2+\left[N(6-N)(s-r)+2s^2\right]\alpha_1^2\alpha_2\beta_2\right\}$$

$$+\Big\{\left[r(20-12N+3N^2)-N(2N^2-10N+16)\right]\alpha_2^2\beta_2+\Big[s(20-12N+3N^2)$$

$$-N(2N^2-10N+16)\Big]\alpha_2\beta_2^2\Big\}\Big]+r\Big[r^2(3N^2-12N+20)-Nr(6N^2-30N+48)$$

$$+N^2(5N^2-24N+32)\Big]\alpha_2^3+s\Big[s^2(3N^2-12N+20)-Ns(6N^2-30N+48)$$

$$+N^2(5N^2-24N+32\Big]\beta_2^3\Big\}.$$

$$s_2(v_4)=\frac{8}{N^8}\Big\{3N^2r(7N^4-20N^3-8N^3r+38N^2r+4N^2r^2-28Nr^2+7r^3)\alpha_1^4\alpha_2^2$$

$$+3N^2s(7N^4-20N^3-8N^3s+38N^2s+4N^2s^2-28Ns^2+7s^3)\beta_1^4\beta_2^2$$

$$+2N^4rs\Big[(s\alpha_1^6\alpha_2+r\beta_1^6\beta_2)+(r\alpha_1^6\beta_2+s\alpha_2\beta_1^6)-2(s\alpha_1^3\alpha_2\beta_1^3$$

$$+r\alpha_1^3\beta_1^3\beta_2)\Big]+3N^2r^2s^2\Big[7(\alpha_1^4\beta_2^2+\alpha_2^2\beta_1^4)-4(\alpha_1\beta_1^3\beta_2^2$$

$$+\alpha_1^3\alpha_2^2\beta_1)+12(\alpha_1^2\alpha_2\beta_1^2\beta_2)\Big]+18N^2rs(s^2\alpha_1^2\alpha_2^2\beta_1^2+r^2\alpha_1^2\beta_1^2\beta_2^2)$$

$$+12N^2rs\left[(N^2-N^2r-2s^2)\alpha_1\alpha_2\beta_1^3\beta_2+(N^2-N^2s-2r^2)\alpha_1^3\alpha\beta_1\beta_2\right]$$

$$+6N^2rs\Big[(5N^2+2N^2r-14Nr+7r^2)\alpha_1^4\alpha_2\beta_2+(5N^2+2N^2s$$

$$-14Ns+7s^2)\alpha_2\beta_1^4\beta_2\Big]+12rs\Big[(-2N^3-3N^2rs+12N^2r$$

$$-15Nr^2+5r^3)\alpha_1\alpha_2^3\beta_1+(-2N^3-3N^2rs+12N^2s-15Ns^2$$

$$+5s^3)\alpha_1\beta_1\beta_2^3\Big]+12N^2rs\Big[s(-N^2+2N-r)\alpha_1\alpha_2^2\beta_1^3+r(-N^2$$

$$+2N-s)\alpha_1^3\beta_1\beta_2^2\Big]+6rs\Big[(3N^2r^2-6N^3+15N^2s-12Ns^2+3s^3$$

$$+8rs^2)\alpha_1^2\beta_2^3+(3N^2s^2-6N^3+15N^2r-12Nr^2+3r^3$$

$$+8r^2s)\alpha_2^3\beta_1^2\Big]+36rs\Big[(-5rs^2-N^4s-N^2r^2+3N^3s-2N^3$$

$$+3N^2r)\alpha_1\alpha_2^2\beta_1\beta_2+(-5r^2s-N^4r-N^2s^2+3N^3r-2N^3$$

$$+3N^2s)\alpha_1\alpha_2\beta_1\beta_2^2\Big]+9rs(9N^3-22N^2-36Nrs+72rs$$

$$+6N^2rs)\alpha_2^2\beta_2^2+18rs\Big[(5r^2+7N^2r^2-14N^3r+N^4r+8N^4$$

$$-29Nr^2+43N^2r-21N^3)\alpha_1^2\alpha_2^2\beta_2+(5s^3+7N^2s^2-14N^3s$$

$$+N^4s+8N^4-29Ns^2+43N^2s-21N^3)\alpha_2\beta_1^2\beta_2^2\Big]+18rs\Big[(N^3$$

$$+4N^2s+5N^2rs+5r^2s-19Nrs)\alpha_1^2\alpha_2\beta_2^2+(N^3+4N^2r$$

$$+5N^2rs+5rs^2-19Nrs)\alpha_2^2\beta_1^2\beta_2\Big]+3r(36r^3-18Nr^3$$

$$+75N^2r^2+3N^2r^3-12N^3r^2-126N^3r+28N^4r+84N^4$$

$$-32N^5+4N^6-126Nr^2+162N^2r-78N^3)\alpha_2^4+3s(36s^3$$

$$-16Ns^3+75N^2s^2+3N^2s^3-12N^3s^2-126N^3s+28N^4s+84N^4$$

$$-32N^5+4N^6-126Ns^2+162N^2s-78N^3)\beta_2^4+6r(5r^4-39Nr^3$$

$$+102N^2r^2+3N^4r^2+9N^2r^3-36N^3r^2-121N^3r+57N^4r$$

$$-6N^5r+60N^4-42N^5+8N^6)\alpha_1^2\alpha_2^3+6s(5s^4-39Ns^3$$

$$+102N^2s^2+3N^4s^2+9N^2s^3-36N^3s^2-121N^3s+57N^4s$$

$$-6N^5s+60N^4-42N^5+8N^6)\beta_1^2\beta_2^3+6rs(72r^2+66N^2-51N^3$$

$$+75N^2r-36Nr^2-126Nr+6N^2r^2+11N^4-12N^3r)\alpha_2^3\beta_2$$

$$+6rs(72s^2+66N^2-51N^3+75N^2s-36Ns^2-126Ns+6N^2s^2$$

$$+11N^4-12N^3s)\alpha_2\beta_2^3\Big\}.$$

$$s_{11}(V_2,V_3)=\frac{6}{N^5}\Big\{r\big[s^2(N-3)+N(N^2-N-s)\big]\alpha_1\alpha_2^2+s\big[r^2(N-3)$$

$$+N(N^2-N-r)\big]\beta_1\beta_2^2+rs\big[s(3-N)\alpha_2^2\beta_1+r(3-N)\alpha_1\beta_2^2\big]$$

$$+rs\big[\{(N-2)(r-s)+2s\}\alpha_1\alpha_2\beta_2+\{(N-2)(s-r)+2r\}\alpha_2\beta_1\beta_2\big]$$

$$+Nrs\big[(s\alpha_1^3\alpha_2+r\beta_1^3\beta_2)-(s\alpha_1^2\alpha_2\beta_1+r\alpha_1\beta_1^2\beta_2)$$

$$-(r\alpha_1^2\beta_1\beta_2+s\alpha_1\alpha_2\beta_1^2)+(r\alpha_1^3\beta_2+s\alpha_2\beta_1^3)\Big\}.$$

$$s_{11}(V_2,V_4)=\frac{4}{N^6}\Big\{3r(13N^2r-7N^3+2N^2r^2+3N^4-4N^3r+2r^3$$

$$-9Nr^2)\alpha_1^2\alpha_2^2+3s(13N^2s-7N^3+2N^2s^2+3N^4-4N^3s$$

$$+2s^3-9Ns^2)\beta_1^2\beta_2^2+3r(6N^2+N^4+3r^2-5N^3-Nr^2+4N^2r$$

$$-8Nr)\alpha_2^3+3s(6N^2+N^4+3s^2-5N^3-Ns^2+4N^2s-8Ns)\beta_2^3$$

$$+6rs(2r^2-4Nr+N^2r+2N^2-N^3)\alpha_1\alpha_2^2\beta_1+6rs(2s^2$$

$$-4Ns+N^2s+2N^2-N^3)\alpha_1\beta_1\beta_2^2+2N^2rs\big[(s\alpha_1^4\alpha_2+r\beta_1^4\beta_2)$$

$$-(s\alpha_1^3\alpha_2\beta_1+r\alpha_1\beta_1^3\beta_2)-(s\alpha_1\alpha_2\beta_1^3+r\alpha_1^3\beta_1\beta_2)$$

$$+(r\alpha_1^4\beta_2+s\alpha_2\beta_1^4)\Big]+3rs\Big[2(4rs-N^3)\alpha_1\alpha_2\beta_1\beta_2+2(2r^2-5Nr$$

$$+2N^2+N^2r)\alpha_1^2\alpha_2\beta_2+2(2s^2-5Ns+2N^2+N^2s)\alpha_2\beta_1^2\beta_2$$

$$+(N^2-2Ns+3rs+s^2)\alpha_1^2\beta_2^2+(N^2-2Nr+3rs+r^2)\alpha_2^2\beta_1^2$$

$$+(9r-8N+4N^2-3Nr)\alpha_2^2\beta_2+(9s\ \ 8N+4N^2-3Ns)\alpha_2\beta_2^2\Big]\Big\}.$$

$$s_{11}(v_3,v_4)=\frac{12}{N^7}\Big\{Nr(6r^3-26Nr+37N^2r+4N^2r^2-8N^3r-20N^3$$

$$+7N^4)\alpha_1^3\alpha_2^2+Ns(6s^3-26Ns^2+37N^2s+4N^2s^2-8N^3s$$

$$-20N^3+7N^4)\beta_1^3\beta_2^2+r(75Nr^2-107N^2r+58N^3-19r^3$$

$$+6Nr^3-30N^2r^2+54N^3r+3N^3r^2-6N^4r-42N^4$$

$$+8N^5)\alpha_1\alpha_2^3+s(75Ns^2-107N^2s+58N^3-19s^3+6Ns^3$$

$$-30N^2s^2+54N^3s+3N^3s^2-6N^4s-42N^4+8N^5)\beta_1\beta_2^3$$

$$+Nrs\Big[(-N^2r+2N^2-2s^2+4rs)\alpha_1^3\beta_2^2+(-N^2s+2N^2-2r^2$$

$$+4rs)\alpha_2^2\beta_1^3\Big]+N^3rs\Big[(s\alpha_1^5\alpha_2+r\beta_1^5\beta_2)-(s\alpha_1^3\alpha_2\beta_1^2$$

$$+r\alpha_1^2\beta_1^3\beta_2)-(r\alpha_1^3\beta_1^2\beta_2+s\alpha_1^2\alpha_2\beta_1^3)+(r\alpha_1^5\beta_2+s\alpha_2\beta_1^5)\Big]$$

$$+Nrs\Big[(-N^3+11N^2+5N^2r-16r^2-28rs)\alpha_1^3\alpha_2\beta_2+(-N^3$$

$$+11N^2+5N^2s-16s^2-28rs)\alpha_2\beta_1^3\beta_2\Big]+3Nrs\Big[(-N^2s$$

$$+3N^2-5Nr+2r^2)\alpha_1\alpha_2^2\beta_1^2+(-N^2r+3N^2-5Ns$$

$$+2s^2)\alpha_1^2\beta_1\beta_2^2\Big]+3Nrs\Big[(-N^2r+N^2-2Ns+4rs)\alpha_1\alpha_2\beta_1^2\beta_2$$

$$+(N^2s+N^2-2Nr+4rs)\alpha_1^2\alpha_2\beta_1\beta_2\Big]+3Nrs\Big[s(N$$

$$-2r)\alpha_1^2\alpha_2^2\beta_1+r(N-2s)\alpha_1\beta_1^2\beta_2^2\Big]+3rs(\ 19r^2+41Nr$$

$$23N^2+6Nr^2-17N^2r+11N^3+2N^3r-N^4)\alpha_1\alpha_2^2\beta_2$$

$$+3rs(-19s^2+41Ns-23N^2+6Ns^2-17N^2s+11N^3$$

$$+2N^3s-N^4)\alpha_2\beta_1\beta_2^2+3rs(-19rs+6Nrs+7Ns$$

$$-N^2-N^3r-N^2s+3N^2r)\alpha_1\alpha_2\beta_2^2+3rs(-19rs+6Nrs$$

$$+7Nr-N^2-N^3s-N^2r+3N^2s)\alpha_2^2\beta_1\beta_2+rs\Big[(3N^3-9N^2s$$

$$+19rs-8Nr+6N^2s)\alpha_1\beta_2^3+(3N^3-9N^2+19rs-8Ns$$

$$+6Nr^2)\alpha_2^3\beta_1\Big]\Big\}.$$